

2006

## The effect of student self -described learning styles within two models of teaching in an introductory data mining course

Matthew A. North  
*West Virginia University*

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

North, Matthew A., "The effect of student self -described learning styles within two models of teaching in an introductory data mining course" (2006). *Graduate Theses, Dissertations, and Problem Reports*. 3471. <https://researchrepository.wvu.edu/etd/3471>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

The Effect of Student Self-described Learning Styles within  
Two Models of Teaching in an Introductory Data Mining Course

Matthew A. North

Dissertation submitted to the  
College of Human Resources and Education  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of

Doctor of Education  
in Technology Education

Terence C. Ahern, Ph.D., Chair  
Sebastian R. Diaz, Ph.D.  
Samuel B. Fee, Ph.D.  
Cheryl L. Prichard, Ed.D.  
R. Neal Shambaugh, Ph.D.

Department of Advanced Educational Studies

Morgantown, West Virginia  
2006

Keywords: Data mining, teaching models, learning styles,  
direct instruction, concept attainment, rules of association

© 2006 Matthew A. North

## ABSTRACT

### The Effect of Student Self-described Learning Styles within Two Models of Teaching in an Introductory Data Mining Course

Matthew A. North

This dissertation examines the roles of learning styles and teaching methodologies within a data mining educational program designed for non-Computer Science undergraduate college students. The experimental design is framed by a discussion of the history and development of data mining and education, as well as a vision for its future.

Data mining is a relatively new discipline which has grown out of the fields of database management and data warehousing, statistics, logic, and decision sciences. Over the course of its approximately 15 year history, data mining has emerged from its genesis within the academic and commercial research and development arenas to become a widely accepted and utilized method of exploratory data analysis for management, strategic planning and decision support. Over the first several years of its development, data mining remained the province of computer scientists and professional statisticians at large corporations and research universities around the world. Beginning in about 1989, these data mining pioneers developed many of data mining's standards and methodologies on large datasets using mainframe computing systems. Throughout the 1990s, as both the hardware and software tools required for the realization of data mining have become increasingly accessible, powerful and affordable, the pool of potential data miners has expanded rapidly. Today, even individuals and small businesses can exploit the power of data mining using freely acquirable open source software packages capable of running on personal computers.

During the growth and development of data mining methodologies however, little research has been dedicated specifically to the pedagogical approaches used in teaching data mining. Educational programs that have evolved have largely remained within Computer Science departments and have often targeted graduate students as an audience. This dissertation seeks to examine the possibility of successful teaching data mining concepts and techniques to a non-Computer Science undergraduate audience. The study approached this research question by delivering a lesson on the data mining topic of Association Rules to 86 participants who are representative of the target audience. These participants were randomly assigned to receive the Association Rules lesson through either a Direct Instruction or a Concept Attainment teaching approach. The students completed Kolb's Learning Styles Inventory, participated in the data mining lesson, and then completed a quiz on the concepts and techniques of Association Rules. A t-test was used to determine if significant differences existed between the scores generated under the two teaching models, and an ANOVA was conducted to identify significant differences between the four learning style groups from Kolb's instrument. In addition to these two statistical tests, the data were also mined using Association Rules and Decision Tree methods.

In both statistical tests, we failed to reject the null hypothesis, finding no significant differences in quiz scores between the two teaching models or among the four learning style groups. Further investigation into the differences among learning styles within teaching models however did reveal that the Assimilator learning style students who received their instruction via Direct Instruction did score significantly higher on the quiz than did their learning style counterparts who received the lesson via Concept Attainment. This finding suggests that although we cannot rely solely on one instructional approach as consistently more effective than the other, there may be instances where the correct instructional choice will positively benefit some learners with certain learning styles. The results of the data mining activities also support this assertion. Association Rules mining yielded no strong relationships between teaching models, learning styles and quiz scores, but Decision Tree mining did reveal a similar pattern of higher scores earned by Assimilator learners within Direct Instruction.

The findings of this study show that effectively teaching data mining concepts to undergraduate non-Computer Science students will not be as simple as choosing one teaching methodology over another or targeting a specific learning style group. Rather, designing instructional activities using teaching methodologies which closely align with predominant learning styles in a classroom should prove more effective. Perhaps the most significant finding of the study is that elementary data mining concepts and techniques can be effectively taught to the target audience. Finally, we recommend that additional teaching methodologies and perhaps different learning style assessments could be tested in the same way as those selected for this study.

## DEDICATION

This dissertation is dedicated to my wife, Joanne, and our three daughters, Jessica, Emily and Madeline, who spent many evenings without daddy at home so that I could complete my doctoral studies. I deeply appreciate their sacrifice, which has enabled my successful completion of this work.

## ACKNOWLEDGMENTS

I would like to first acknowledge my parents, Larry and Roberta North, who taught me from an early age to finish what I start, work hard, live an honest life, and accept accountability for myself and my actions. They also reinforced throughout my life that failure to finish college was unacceptable, which instilled in me the desire necessary to pursue and complete this doctoral degree.

I would also like to acknowledge Joanne's parents, Willis and Dianne Fails, who have consistently supported the academic pursuits of both Joanne and me since our marriage 11 years ago. I am grateful for their example and encouragement.

In my formative years, two school teachers stand out from the rest in my mind. My 5<sup>th</sup> grade teacher, Mrs. Janelle Stockman, and my 8<sup>th</sup> grade teacher Ms. Janice Quaney helped me to build knowledge, but also to develop the self confidence needed to successfully complete graduate work. I thank them for the dedicated educators they were for me.

Finally, I would like to thank my committee members, Drs. David McCrory, Terence Ahern, Cheryl Prichard, Sebastian Diaz, Neal Shambaugh and Samuel Fee, with special thanks to Drs. McCrory and Ahern for their guidance as my advisors during my doctoral studies. I offer my sincere thanks to each of these mentors for their positive reinforcement, guidance, and when necessary, correction.

## TABLE OF CONTENTS

ABSTRACT .....	ii
DEDICATION .....	iv
ACKNOWLEDGMENTS .....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
CHAPTER I: INTRODUCTION.....	1
A Brief History of Data Mining.....	2
<i>Early Contributors</i> .....	2
<i>Corporate Influence</i> .....	3
<i>Academic Offerings</i> .....	4
Data Mining for the Masses.....	5
Significance of the Study .....	6
Examining Learning Styles.....	7
Examining Models of Teaching.....	8
Statement of the Problem.....	9
Research Questions.....	10
Operational Definitions.....	11
Limitations .....	14
Summary .....	15
CHAPTER II: REVIEW OF LITERATURE .....	17
Data Mining Research.....	17
<i>Historical Timeline</i> .....	17
<i>An Emerging Interdisciplinary Field</i> .....	21

Learning Styles Research.....	25
<i>Definitions, History and Development</i> .....	26
<i>Kolb and Categorization</i> .....	29
<i>Online Learning</i> .....	31
<i>Summary of Learning Styles Literature Review</i> .....	33
Models of Teaching .....	34
<i>Review of Models of Teaching</i> .....	34
<i>Mapping Models of Teaching to Learning Styles</i> .....	37
<i>Selecting Models for this Study</i> .....	39
<i>Direct Instruction</i> .....	40
<i>Concept Attainment</i> .....	41
<i>Online Teaching</i> .....	41
<i>Summary of Teaching Model Literature Review</i> .....	42
The Pedagogy of Data Mining.....	42
Summary of Literature Reviewed.....	43
CHAPTER III: METHOD AND PROCEDURE.....	44
Organizational Model .....	47
Participants.....	47
Instruments.....	48
<i>Study Overview</i> .....	48
<i>Demographics</i> .....	49
<i>Learning Styles</i> .....	50
<i>Rules of Association via Direct Instruction</i> .....	51
<i>Rules of Association via Concept Attainment</i> .....	53
Data Collection .....	54



Data Analysis .....	54
<i>Demographic Analysis</i> .....	55
<i>Analysis of Models of Teaching</i> .....	55
<i>Learning Style Analysis</i> .....	55
<i>Level Two Hypotheses</i> .....	56
<i>Summary of Research Methodology</i> .....	57
CHAPTER IV: RESULTS.....	58
Outcomes of Data Analysis .....	58
<i>Demographic Data and Discussion of Participant Profiles</i> .....	58
<i>Findings for Research Question One: Models of Teaching</i> .....	61
<i>Findings for Research Question Two: Learning Styles</i> .....	64
<i>Experiment-wise Error Rate</i> .....	69
<i>A Data Mining Approach to the Data</i> .....	69
<i>Summary of Study Findings</i> .....	75
CHAPTER V: SUMMARY, DISCUSSION AND RECOMMENDATIONS.....	76
Summary and Discussion.....	76
<i>Discussion of Demographic Data</i> .....	77
<i>Outcomes for Models of Teaching</i> .....	79
<i>Learning Style Outcomes</i> .....	81
<i>Other Outcomes of Data Analysis</i> .....	85
Recommendations.....	87
<i>Implications for Data Mining Education</i> .....	87
<i>Future Study</i> .....	89
Conclusion .....	91
REFERENCES .....	93

APPENDICIES .....	106
<i>Appendix A: IRB Approval Forms</i> .....	106
<i>Appendix B: Kolb’s Learning Styles Inventory (Kolb, 1985)</i> .....	109
<i>Appendix C: Direct Instruction Assessment Instrument</i> .....	110
<i>Appendix D: Code for Online Instruction and Data Collection</i> .....	113
<i>Appendix E: Complete Participant Data Set</i> .....	127

## LIST OF TABLES

Table 2.1. <i>Teaching Model Families</i> .....	36
Table 4.1. <i>Summary of Data Collected</i> .....	59
Table 4.2. <i>Descriptive Statistics for Concept Attainment Participants</i> .....	62
Table 4.3. <i>Descriptive Statistics for Direct Instruction Participants</i> .....	63
Table 4.4. <i>Descriptive Statistics for Quiz Scores within the Accomodator Learning Style</i> .....	65
Table 4.5. <i>Descriptive Statistics for Quiz Scores within the Assimilator Learning Style</i> .....	66
Table 4.6. <i>Descriptive Statistics for Quiz Scores within the Converger Learning Style</i> .....	67
Table 4.7. <i>Descriptive Statistics for Quiz Scores within the Diverger Learning Style</i> .....	68
Table 4.8. <i>Decision Tree Rule Exemplars from the Study Data</i> .....	72
Table 5.1. <i>Quiz Score Means by Learning Style and Teaching Model</i> .....	86
Table D.1. <i>Complete Participant Data Set</i> .....	127

## LIST OF FIGURES

Figure 2.1. Kolb's Learning Style Categories .....	30
Figure 3.1. Dissertation Hypothesis Map .....	46
Figure 3.2. Organizational Model .....	47
Figure 3.3. Study Overview and Participant ID Registration .....	49
Figure 3.4. Demographic Data Collection Instrument .....	50
Figure 3.5. Kolb's Learning Styles Inventory .....	51
Figure 3.6. Direct Instruction Learning Module for Association Rules in Data Mining .....	52
Figure 3.7. Assessment Instrument (Quiz) for the Association Rules Learning Module .....	53
Figure 3.8. Concept Attainment Learning Module for Association Rules in Data Mining .....	54
Figure 4.1. Frequency Distributions for Quiz Scores within Models of Teaching .....	64
Figure 4.2. Decision Tree Exemplar from the Study Data .....	71
Figure 5.1. Participant group sample sizes by Learning Style .....	82
Figure 5.2. Quiz Score Frequency Distributions by Learning Style .....	83
Figure 5.3. Quiz Score Means, Medians and Modes by Learning Style .....	84

## CHAPTER I: INTRODUCTION

Data Mining is an emerging field in information management and analysis. In its limited history, it has emerged as a powerful component of business intelligence, supporting efforts to enhance decision making, increase customer satisfaction, disseminate information, predict and prevent fraud, and improve marketing and promotion efforts. As success with the technology has grown, so too have desires by both corporate and non-profit organizations to exploit this new tool to solve modern business problems or identify opportunities for growth (Han & Kamber, 2001).

The academic community has aided the expansion of data mining technology as colleges and universities world-wide have contributed to the definition and extension of data mining as a discipline, used data mining as a tool in their own operations, and developed educational programs aimed at contributing to the understanding, uses and enhancement of data mining (Luan, 2002).

As educational programs in data mining have emerged and improved, a strong trend has occurred where the majority of related course offerings exist either in upper-division undergraduate Computer Science programs, or in graduate degree programs in Computer Science or other quantitative professional programs in business analytics or statistics (Urbancic, Skjanc & Flach, 2002). This is understandable given the historically strong reliance of data mining processes on high technology—data warehouses measured in terabytes, statistical analysis packages capable of handling millions of rows of data in a single operation, servers possessing multiple processors, random access memory measured in gigabytes, and so on. However, the basic concepts and techniques of data mining do not, by nature, demand such a grand scale in order to be useful to organizations (Waldrop, 2001). It is the belief in this assertion which fuels

the preparation of this dissertation: Data mining *can* be taught to undergraduate college students studying in *any* field, in ways they can understand, preparing them to apply data mining as a tool regardless of their future career choice or the size of their future employer's information technology budget. This study will examine the relationship between the self-described learning styles of undergraduate college students not majoring in Computer Science with two intentionally selected models of teaching which are highly complementary to the instructional demands of teaching data mining.

### A Brief History of Data Mining

In order to address the overarching goal of this study, it is important to understand how data mining emerged. As with many technologies throughout history, data mining has evolved from a series of problem solving and solution creation efforts, forming a loose collection of tools with few standards for compatibility, interface design or output (Fayyad & Uthurusamy, 1996). Due to its strong ties with data management tools such as relational database management systems, data warehouses, statistical analysis software, and artificial intelligence, it logically follows that data mining grew out of, and remains closely tied to, experts in computer and software engineering, psychology, and statistics.

#### *Early Contributors*

In 1989, Piatesky-Shapiro presented a workshop at the *International Joint Conferences on Artificial Intelligence* (IJCAI) titled Workshop on Knowledge Discovery in Databases. Two years later, a book by the same title was released, along with a flurry of additional workshops at academic conferences in both Computer Science and artificial intelligence. Decreases in the cost of increasingly powerful computing technology in the mid-1990s fueled greater interest in data

mining, which in turn served as a catalyst for more widespread adoption of data mining techniques as well as increased research in the field. In 1995, the first major conferences dedicated specifically to knowledge and data discovery were held, subsequently leading to the creation of the *Journal of Data Mining and Knowledge Discovery* and the formation of the Knowledge and Data Discovery Special Interest Group (SIG-KDD) of the Association for Computing Machinery (ACM), for years the foremost professional and academic association in the computing sciences. By 1998, the field of data mining was becoming increasingly recognized by most academics and professionals working the fields where data mining first had its genesis (Piatetsky-Shapiro, 1999). SIG-KDD founded their own conference in 1999, adding needed organizational activities and forums for dialogue regarding standards to the fledgling discipline.

### *Corporate Influence*

The relationship between academe and the corporation is often recursive (Agrawal, 2003; Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro & Simoudis, 1996). For example, statistics software giant SAS was founded by Dr. James Goodnight, who identified a great business need for the types of statistical analysis software he was developing as a postgraduate scholar at North Carolina State University (Zaiane, 2006). Since founding SAS, Goodnight has branched out into providing data mining tools for both education and corporate enterprises. Competitors have followed suit. SPSS, Inc. has released Clementine, its own data mining package, and other segment players including Oracle and IBM have added their own data mining software offerings. Hardware manufacturers have anxiously contributed to the growth of data mining by providing increasingly powerful computing devices which support storage and processing of vast amounts of data (Exner & Bear, 1998). Sun Microsystems, Hewlett-Packard, Dell Computer, NCR and

IBM have all created divisions dedicated to creating, configuring and selling hardware equipment designed to process data mining tasks.

### *Academic Offerings*

Since academicians at large research universities across the globe were among the first to embrace and contribute to the definition, growth and capabilities of data mining, it makes sense that the first course offerings on these topics followed, or accompanied, the research being conducted in the field (Frawley, Piatetsky-Shapiro & Matheus, 1992). As a natural consequence of this phenomenon, the courses were offered in the areas where the researchers worked: Computer Science, Engineering, and Artificial Intelligence – primarily at the graduate level. Early programs at the University of Illinois—Urbana/Champaign, Simon Fraser University, and Massachusetts Institute of Technology paved the way for a rapid expansion of course offerings not only in North America, but also in Europe and Asia (Fayyad, Piatetsky-Shapiro & Smyth, 1996). By 2001, entire graduate programs of study were offered, often with a niche in some market segment where data mining represented a potentially useful tool—data mining for health care, finance, or customer relationship management, for example (Urbancic, et al, 2002; Harper, 2005).

In the midst of these internationally recognizable university names and marquee corporate logos however, there is a different market for data mining education, which has received relatively little attention in the first decade of data mining development (Han, Hu & Cercone, 2003). Small and mid-sized companies with smaller computing budgets and fewer resources for hiring full-time graduate or post-graduate statisticians, computer scientists, or logicians stand to reap benefits from data mining activities as well.



## Data Mining for the Masses

Today, data mining has begun to show signs of extending toward this group (Han, Altman, Kumar, Mannila & Pregibon, 2002; Ji, Liu, Sha, & Zhong, 2005). Recently released tools, including some open source data mining software (OSS) packages, incorporate graphical user interfaces (GUIs) and natural, intuitive interfaces with more familiar analysis software packages such as Microsoft Excel (Chung & Cheung, 2005). OSS is specifically mentioned here for two reasons. First, data mining software has traditionally been expensive and difficult to obtain, requiring consultation with vendors and extensive contracts not only for acquisition, but also for installation, configuration and support (Potvin, 2004). OSS on the other hand, is provided free under the GNU Public License (GPL), enabling a new segment of the population access to the types of software tools previously limited to large corporate customers (Stallman, 2004). Additionally, OSS spurs competition in software markets, driving developers of both free and proprietary systems to improve features and stability, compete for customers, and expand awareness of the discipline.

Whether open source or not, this next generation of tools provides a platform and climate which will enable a broader spectrum of educators to reach non-Computer Science undergraduate college students with course offerings in data mining (Chung & Cheung, 2005). Determining the most appropriate ways to teach data mining concepts and techniques to the audience in meaningful and effective ways is a primary step toward effective expansion of course offerings beyond the traditional bounds set during the first ten to twelve years of data mining education (Urbancic, et al. 2002). If successful, this new type of data mining student may eventually show, through effective use of data mining within a limited computing infrastructure and in a smaller work environment, that data mining course offerings to this target

group is not only appropriate, but beneficial. That cannot be known however, without first addressing the need to deliver data mining education to this new student population segment in accessible, useful ways given their background, experience, and aptitudes.

### Significance of the Study

Currently there exists a relative paucity of scholarly research in the area of data mining education (Urbancic, et al., 2002). Most of the research conducted and published in the approximately 15 years of data mining's existence has been about the techniques, concepts and algorithms used to *conduct* data mining (Piatetsky-Shapiro, 1999; Rao, 2001). Little attention has been paid to the effectiveness of *teaching* techniques used to impart data mining skills and knowledge to students. A clear understanding of how students' learning styles impact their abilities to acquire data mining knowledge and skills will help to shape the future of data mining education and better prepare college graduates for an increasing number of jobs in this field.

College and university faculty members are constantly reviewing and revising curriculum in order to ensure that it is recent, relevant and practically applicable (Dringus & Ellis, 2005). This is particularly true in the fast-evolving fields of Computer Science and Information Systems and Technology. An understanding of the interaction between student learning styles and models of teaching within the context of data mining courses will help faculty ensure that teaching of these new and highly useful concepts and techniques is effective (Mupinga, Nora & Yaw, 2006). A delivery of course content that is consistent with students' aptitudes and prior knowledge will be more likely to contribute to a positive student outcome; and an instructor who understands these influences on student learning will be able to work within these bounds to teach data mining meaningfully to their audience (Thompson, 1997). By recognizing how one prefers to learn, and then by receiving instruction which is well suited to that learning style, the

probability of acquiring and retaining data mining knowledge and skills is increased (Claxton & Murrell, 1987).

Employers of data miners will also appreciate improvements in data mining education which may result from this study. Employees are generally the largest expense an organization has, and employers hope to get the most for their money from each member of the organization (Rygielski, Wang & Yen, 2002). Those hired after having received data mining instruction in the context of their own learning style will be better prepared to meet the demands of their job.

Beyond the scope of data mining instruction however, this study stands to benefit teaching in general and online education as well. This is because the structure of the research can be universally applied to any content or curriculum (Claxton & Murrell, 1987; Reichmann, 1978). The questions posed in this dissertation are about the interaction and effect of learning styles within different models of teaching. The content here is data mining, however the same research methodology could be applied to art history, physics, or political science. By understanding the relationships among learning styles, teaching methods and curricular content, teachers can create increasingly effective instructional environments to the benefit of all involved.

### Examining Learning Styles

Varying opinions on the usefulness and meaning of learning style assessments exist within the educational community (Sharp, 1997; Thompson, 1997). Functionally, an assessment of one's learning style is an effort to determine the ways in which a student best acquires new knowledge—or at least the ways in which the student *believes* they best acquire new knowledge (Kolb, 1976). Because the participants in this study will come from a wide variety of backgrounds and have an equally diverse set of educational goals, it is important to understand

how each participant feels and thinks about the ways they learn. This information will be useful in interpreting our findings, by allowing us to search for connections between certain learning styles and specific performance outcomes.

For this study, Kolb's Learning Style Inventory is used, since it has a well established track record in educational research and is not designed toward any specific intellectual predisposition (Kayes, 2005; Kolb, 1985). It is well suited to this particular experiment because it attempts to assess both *how the person learns* and *how the person deals with ideas*. Data mining demands both technical and theoretical understanding and ability (Tan, Steinbach & Kumar, 2005), so the option to evaluate learning *and* idea processing mechanisms is beneficial in the context of the teaching and learning to be examined. Learning data mining also demands a high degree of interaction with computer technology, and Kolb's inventory helps to identify individuals who strongly gravitate toward, or away from, learning through the use of hands-on interaction and activity (Searson & Dunn, 2001). Once the self-described learning style of each study participant is identified, we are to determine if their predisposition toward learning impacts their ability to acquire knowledge about data mining concepts and techniques.

### Examining Models of Teaching

Aside from evaluating the ways which students tend to learn best, there remains a question regarding the appropriate delivery mechanism for teaching data mining to the target audience. Here we examine two specific models of teaching defined by Joyce, Weil and Calhoun (2000)—Direct Instruction and Concept Attainment.

Direct Instruction is the process of teaching which presents information as a factual body of knowledge to be acquired by the learner (Joyce, et al., 2000). It is, by nature, a behavioral model in which repetition, planned practice, and instructor feedback shape the learner's

knowledge acquisition process. The instructor, texts, exercises and other learning material all fulfill the role of guiding the learner toward specifically predefined (or “correct”) outcomes. It has emerged as a popular instructional model, particularly in quantitative disciplines such as mathematics where arrival at “correct” solutions is more easily defined (Allen, 1996). Given data mining’s heavy reliance on quantitative methods, this instructional model is a logical choice for this experiment.

The second teaching approach used is Concept Attainment. This model is much more cognitive and even somewhat constructive in nature, allowing the student to learn through a process of guided discovery, creating associations of new information or knowledge with previously understood content (Joyce, et al., 2000; Jonassen, Bessner & Yacci; 1993). As previously mentioned, data mining owes at least part of its genesis to mankind’s attempts to model human cognition through artificial intelligence. Many of the important concepts and techniques encompassed in data mining rely upon computer simulations and algorithms defined specifically to mimic human decision making patterns and mechanisms (Ying, Murphy & Ng, 2004). Thus, there is a strong rationale for including Concept Attainment in this study—the very teaching model is an exemplar of the data mining content to be learned.

#### Statement of the Problem

The purpose of this study is to identify possible relationships or interactions of student-described learning styles within the two selected models of teaching in the context of data mining education. The specific educational experience is targeted to an audience of non-Computer Science undergraduate college students—an audience quite different from the traditional data mining student group. An understanding of how the target audience’s learning styles affect their learning of data mining concepts and techniques within the chosen teaching

approaches is needed to improve the pedagogy of data mining as a whole. The findings of this study, while targeted toward the teaching of data mining, are extensible to the study of learning style/teaching model interactions across a variety of educational topics.

The specific variables to be examined include the following:

1. **Dependent Variable:** The dependent variable in this research is the assessment score collected for each participant at the end of the learning module. Assessment scores are the indicators of successful knowledge or skill acquisition, and were collected and associated specifically with the model of teaching used in delivery—either Direct Instruction or Concept Attainment.
2. **Independent Variables:** The independent variables in this study are the students' self-described learning styles from Kolb's Inventory, and the Direct Instruction and Concept Attainment models of teaching.

Demographic data such as major area of study, age, and sex were also collected and evaluated for possible influence on the student's performance scores, however these were collected and used as explanatory mechanisms and not specifically as independent variables used in addressing the major hypotheses of the study.

### Research Questions

The study is designed to address the following hypotheses:

1.  $H_0$ : There is no significant difference in assessment scores when either Direct Instruction or Concept Attainment is used as the model of teaching in undergraduate, non-Computer Science data mining instruction.

- a.  $H_a$ : There is a significant difference in assessment scores when Direct Instruction or Concept Attainment is used as the model of teaching in undergraduate, non-Computer Science data mining instruction.
2.  $H_0$ : There is no significant difference in assessment scores among various learning styles in undergraduate, non-Computer Science data mining instruction.
  - a.  $H_a$ : There is a significant difference in assessment scores for students categorized in certain learning styles in undergraduate, non-Computer Science data mining instruction.
3.  $H_0$ : If in Research Question #1 above,  $H_a$  is true, there is no significant difference in assessment scores associated with students' learning styles within the model of teaching which yielded the significantly higher scores.
  - a.  $H_a$ : If in Research Question #1 above,  $H_a$  is true, there is a significant difference in assessment scores associated with students' learning styles within the model of teaching which yielded the significantly higher scores.
4.  $H_0$ : If in Research Question #2 above,  $H_a$  is true, there is no significant difference in assessment scores between the two models of teaching which can be associated with students' learning styles.
  - a.  $H_a$ : If in Research Question #2 above,  $H_a$  is true, there is a significant difference in assessment scores between the two models of teaching which can be associated with students' learning styles.

### Operational Definitions

The following operational definitions are used throughout this study:

1. Learner: The individual seeking to acquire knowledge through study, instruction, and social interaction. For this study, all learners are undergraduate college students not majoring in Computer Science.
2. Teacher: The individual charged with the arrangement and facilitation of the learning experience. For this study, the instructor prepared the online learning modules and assessments, recruited participants and provided instructions to them, and evaluated the outcomes of the learning assessments.
3. Learning Style: The process and preference of a learner for perceiving and processing information in order to understand, remember, and solve problems in order to acquire new knowledge and skills in a specific subject matter (Kolb, 1976; 1985).
4. Model of Teaching: The intentional arrangement of a learning environment through preparing and organizing instructional materials, delivery mechanisms, and assessment instruments (Joyce, Weil & Calhoun, 2000).
5. Direct Instruction: A pattern of teaching that consists of the teacher's explanation of a new concept or skill to a group of students, having them test their understanding by practicing under teacher direction, and encouraging them to continue practice (Joyce, Weil & Calhoun, 2000).
6. Concept Attainment: The search for and listing of attributes that can be used to distinguish exemplars from nonexemplars of various categories (Bruner, Goodnow, & Austin, 1967).
7. Data Mining: "The task of discovering interesting patterns from large amounts of data where the data can be stored in databases, data warehouses or other information repositories. It is a young, interdisciplinary field, drawing from areas including database



systems, data warehousing, statistics, machine learning, data visualization, information retrieval, and high-performance computing” (Han & Kamber, 2001).

8. Association Rules: The systematic identification of combinations of variable (column) values that frequently occur in an observation (record), which are found to meet both a minimum support and a minimum confidence threshold. For example, consider a data set of 1,000 grocery store receipts (observations). If the support threshold is 5% and the confidence threshold is 60%, at least 50 of the 1,000 receipts must contain a given combination of items purchased, and for all receipts which contain one of the items identified in a given combination, at least 60% of those receipts must have the other item (or items) in the combination (Han & Kamber, 2001; Soukup & Davidson, 2002).
9. Online Learning: The process of acquiring or enhancing knowledge or skills, facilitated by Internet technologies such as the World Wide Web, eMail, Chat, Discussion Forums, etc. (Dabbagh & Bannan-Ritland, 2005; Minoli, 1996; Simonson, Smaldino, Albright & Zvacek, 2003).
10. Assessment Score: The numerical representation of how well each student performed on the quiz (assessment) administered at the end of the learning module. The assessment determines the learners’ cognitive understanding of Rules of Association after content delivery by way of either the Direct Instruction or the Concept Attainment model of teaching. The assessment represents each study participant’s performance as a number of questions correct out of ten.
11. Study Participant: Whenever study participants are referred to, these are individual undergraduate college students, not majoring in Computer Science, at a traditional, residential teaching college.

## Limitations

The following limitations to this study are recognized by the author:

1. The study was conducted at a small, private, coeducational Liberal Arts college in the eastern United States. While the participants are certainly a high interest group for this study, they do not necessarily represent the common undergraduate, non-Computer Science student nationwide. Therefore, generalizability to other institutions is somewhat compromised, more so for schools which may have a higher concentration of non-traditional undergraduate students.
2. The collection of student learning styles is dependent upon each study participant. While it is expected that students will answer the questions on the Learning Styles Inventory honestly, there is no way to validate this data. In order to mitigate this limitation, Kolb's inventory, which is widely accepted in academic research on learning styles, has been selected. We also mitigate this limitation by referring to "student self-described learning styles".
3. The instructional events for this study will take place on the Internet. A general assumption is made regarding study participants' abilities to use a web browser and interact with web-based content—forms, buttons, radio and check boxes, dropdown menus, etc. This limitation will be partially mitigated by asking the student to self-describe their Internet comfort-level and usage habits while collecting demographic data.
4. Data collection for this study will be limited to a one-time-only event during the year 2006. Seasonal factors such as end-of-semester events (e.g. impending holidays, graduation, final exams, job searches, etc.) may have influence the participants' willingness to spend time on, or participate at all, in the study.

## Summary

Data mining as a discipline has reached a level of maturity today which justifies an examination of effective teaching of its core concepts and techniques. Since its inception about 15 years ago, the concepts and techniques within data mining have become both more *defined* and *refined*. Standards have been developed to introduce consistency into the field, and an increase in computing power coupled with a decrease in computing costs has made the use of data mining realistic for an ever-growing audience. Data mining software is now accessible to most people, often with graphical user interfaces which improve usability for non-computer scientists who may lack the programming and algorithmic background previously required in order to conduct data mining tasks. With this shift in technologies in the field, many individuals and organizations across disciplines such as customer service, marketing, risk analytics, health care, etc., are looking to harness the power of data mining in order to enhance their efforts and decision making abilities. It is therefore logical that we examine the ways we are teaching data mining, with an eye toward improving pedagogical performance and student outcomes for a broader student audience.

Although it has been primarily the province of graduate and Computer Science programs for the bulk of its history, it is useful to experiment on the idea of providing data mining instruction specifically to undergraduate college students not studying Computer Science—given current employment conditions, it is in fact a timely decision to consider such curricular offerings. As mentioned, many modern data mining tools no longer require complex programming language skills in order to successfully conduct data mining activities. Thus, an exploration of data mining education at lower academic levels and outside the traditional academic home for these courses seems prudent. Educational programs in business and social

sciences may benefit from offering data mining related coursework, however unlike Computer Science, these academic departments tend to attract a broader range of students, who come to the table with varied learning styles and aptitudes. It is therefore more urgent that teachers in these areas who consider offering data mining education consider their delivery and content in light of these student attributes.

Preparing coursework for the target audience for this study can be done more appropriately by understanding first the learning styles of the audience members, so that instructional materials may be prepared which will effectively facilitate data mining knowledge and skill acquisition. The objective of this dissertation is therefore to evaluate the relationship between student learning styles within different teaching models in undergraduate, non-Computer Science instruction. Knowing this can help teachers prepare to most effectively educate data mining students who fall within these parameters.

## CHAPTER II: REVIEW OF LITERATURE

This study seeks to address three main components: 1) data mining education; 2) student learning styles; and 3) models of teaching. Through an examination of the relationships between student learning styles and models of teaching, we hope to improve data mining education in general. The following sections highlight where we have come from, where we are now, and what appears to be on the horizon, within each of the three component parts of this study.

### Data Mining Research

Data mining experts recognize the genesis of the field beginning between about 1989 and 1991 (Exner & Bear, 1998; Han, Altman, Kumar, Mannila, & Pregibon, 2002; Piatetsky-Shapiro, 1991). Therefore, at the time of this writing the discipline is less than 20 years old. In spite of its brief history, data mining has achieved recognition as a legitimate discipline with a promising future (Piatetsky-Shapiro, 1999). Demand for knowledge and skill in data mining continues to increase across a broad range of disciplines (Jones, 2002; Rygielski, Wang, & Yen, 2002; Saygin & Ulusory, 2002; Shi, Peng, Xu, & Tang, 2002). As demand drives continued evolution and adoption of data mining, educational research must also evolve in order to ensure that data mining instruction remains both consistent and concurrent with actual practices and techniques (Rao, 2001). Educational research in data mining remains a small piece of the overall research conducted in the field to date.

### *Historical Timeline*

Data mining was pioneered by statisticians, logicians, and computer scientists at some of the world's most renowned universities and corporations (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro, & Simoudis, 1996). As early as the mid-1980s, the foundational technologies

and algorithms which would eventually enable data mining were beginning to emerge. For example, beginning in 1987, data mining-type languages were being developed for use in large databases by researchers at GlobeTel Communications Corporation (GTE) (Piatetsky-Shapiro & Jacobson, 1987). These early efforts in the area of data mining required innovative thinking and a fusion of concepts and techniques from the fields of Logic, Statistics, and Database Management. This prompted partnerships with academic researchers, which resulted in the first data mining methodologies and models (Jakobson, Lafond, Piatetsky-Shapiro, & Rajinikanth, 1988). The CALIDA model, which emerged from the aforementioned study, proposed a method for aggregating and then analyzing data from multiple heterogeneous databases (e.g. marketing, customer service, production, etc.). Though rudimentary by today's computing standards, the theoretical underpinnings of the CALIDA model remain as relevant now as they were revolutionary when first developed.

The early 1990s brought much excitement to the world of data mining, accompanied by a broad range of interested and talented individuals in related fields (Frawley, Piatetsky-Shapiro, & Matheus, 1992; Piatetsky-Shapiro, 1991; Piatetsky-Shapiro, Matheus, & Smyth, 1994). Piatetsky-Shapiro's 1991 publication is a defining document for the discipline, helping professionals for the various contributing disciplines of data mining to differentiate and identify what actually characterized the field of Data Mining. Many of these principal components were further defined, and refined, by Frawley, et al. in 1992. In this study, researchers attempted to create a conceptual road map for data mining, characterized by their own activities in the field along with the work of others contributing to early developmental phases. This framework further identified the scope and goals of data mining and encouraged further interest from a growing number of academics and professionals in information management and analytics. Only

a few years later, Piatetsky-Shapiro, Matheus, and Smyth, (1994) reported a sort of state-of-the-union, reflecting on the formally recognized professional conference on Knowledge and Data Discovery in databases, which had been held in late 1993. This publication highlighted progress in data mining in terms of tools and techniques, which had vastly improved as a result of advancements in high-performance computing hardware and software. Perhaps most critical from Piatetsky-Shapiro, et al. however was a recognition and documentation of the challenges and problems which faced data mining at that time. This codification of outstanding issues created a research agenda platform for aspiring researchers in the field.

Throughout the mid-1990s, rapid advancement toward solutions to these problems characterized data mining research. Brachman, et al. (1996), tackled several problems specifically facing business consumers of data mining by evaluating various potential solutions to common business problems which complicated effective data mining by these organizations. Their findings helped to standardize many of the principal data mining technologies used today, including Rules of Association (Induction Rules), Decision Trees, and k-Means clustering. The standardization of methods and procedures helped to introduce a needed degree of stability to the fledgling field (Piatetsky-Shapiro, 1999).

Two additional studies emerged in 1996 from Fayyad, Piatetsky-Shapiro, and Smyth which targeted the contemporary challenges in the field. The first, published in the periodical *Artificial Intelligence*, attempted to frame data mining in the context of its outcomes by reminding the community that the original objectives were to realize “Knowledge Discovery in Databases (KDD).” These researchers felt strongly that data mining users needed to renew their focus on correct and appropriate interpretations of data mining results for specific organizational purposes, rather than the processes themselves. These researchers followed up with a subsequent

study in *Communications of the ACM*, which proposed “The KDD Process for Extracting Useful Knowledge from Volumes of Data.” This effort facilitated further standardization while naturally shifting data miners’ attention away from the process itself and toward the eventual analysis of outcomes of data mining activities—or in other words, toward the knowledge which can be gained through data mining. A third 1996 study (Fayyad & Uruthusamy) provided additional clarification and focus on analysis and responsible usage of data mining results.

Further formative progress in data mining was realized in the latter years of the 1990s. The Cross Industry Standard Platform for Data Mining (CRISP-DM) resulted from collaborative efforts of corporations, software providers and researchers in the field between about 1996 and the release of version 1.0 of the methodology in 1999 (Chapman, et al., 1999). Exner and Bear (1998) hailed this type of work as critical to the long term viability of data mining while highlighting four years of progress on many of the pressing issues exposed in the aforementioned 1994 status report from Piatetsky-Shapiro, Matheus and Smyth. Exner and Bear recognized in their publication the importance of emerging standardization of terminology, methodology, and interpretation. The resulting consistency within the discipline encouraged adoption and spurred additional growth in both implementation and research activity (Gray, 1997). The cumulative effect of the final five years of the 1990s led Piatetsky-Shapiro, who by then had forged a reputation as one of the preeminent figures in data mining and knowledge discovery, to proclaim “the coming of age of the data mining industry” (1999). Although many celebrated this so called coming of age, research continued to underscore the need to fill gaps and provide additional validation of data mining concepts and techniques. Within weeks of Piatetsky-Shapiro’s statement, Bradley, Fayyad, and Mangasarian (1999) followed with a publication of scholarly work on new formulations in data mining, which they found addresses some previous needs



while at the same time uncovering as yet undiscovered degrees of complexity in data mining programming. Surprisingly, even with all of the advancements in both hardware and software during the previous decade, mathematical computing power was found to be insufficient for some of the most promising approaches to data mining. The authors concluded that finding solutions to mathematical limitations in computing power posed a major hurdle for data mining. Thus the stage was set for the next phase of data mining history as the new millennium began.

### *An Emerging Interdisciplinary Field*

The catalysts for data mining are diverse. Even during the formative years of the discipline, rarely was any project undertaken solely for the purpose of furthering data mining—rather, each objective was to further data mining *for some purpose* (Gray, 1997). The financial sector looked to data mining to provide risk scores and return-on-investment modeling; the customer service arena sought improved understanding of consumer habits and expectations in order to drive retention; while marketing professionals hoped to expand reach while improving target deliveries (Efthimiadis, 2000). Even areas not traditionally associated with the quantitative approaches used in data mining looked for, and sometimes found, useful purposes for data mining within their individual sphere. Agrawal (2003) and Anane (2001) both published examples which illustrate the broad spectrum of applications researchers have begun to find for data mining techniques. In these papers, the authors specifically discuss the usefulness of data mining in the fields of humanities and social sciences. While not immediately evident, these and other researchers have uncovered additional knowledge about literature, authorship, and attitudes through techniques such as text, visual and audio data mining—all extensions of the traditional approaches to data mining which were formalized through publication shortly after the start of the new millennium (Han & Kamber, 2001).

In more traditionally quantitative fields, data mining also gained attention and began to branch out at the turn of the new century. A 2000 study introduced a simplified yet powerful algorithm for identifying and measuring association strengths between variable observations in a data set (Cheung, Hwang, Fu, & Han). Further research specifically in the area of association rules followed, helping to detail the uses and functions of this type of data mining (Tung, Hongjun, Han, & Feng, 2003). Rules of association have subsequently become a staple data mining tool available today.

Chiu (2004) built from this platform by extending the rule-based induction approach of Cheung, et al. to the World Wide Web. In the period of dot com proliferation leading up to the year 2000, online marketing, sales and competition reached a fevered pitch, often with investors blindly funding poorly planned and flimsy enterprises. Business plans and strategies were seen as time consuming paperwork in the new fast-paced online environment. After the so-called Internet bubble burst after the New Year, investors were left reeling, businesses were shut down, and reality set in. New ventures would need to be founded in realistic research and actual thought—backed up by sound business planning and methodologies. Data mining could be applied in this environment to track, organize and evaluate online behaviors, enabling would be eCommerce players to make intelligent and informed decisions. According to Chiu, not only could the online environment be mined in order to provide this data, but it could also be used to deliver it. Even individual privacy acquired a new level of attention in this more methodical approach to data mining. While association rules based on Cheung, et al.'s work continued to gain popularity as a powerful data mining tool, attention was turned to employing this and other data mining techniques in responsible and secure ways, to the benefit of consumers and organizations (Evfimievski, Srikant, Agrawal, & Gehrke, 2004).

The maturity in the field brought about by this sobering series of events following the advent of the new century also helped to solidify some of the more pragmatic aspects of the discipline of data mining. As previously mentioned, standards were beginning to emerge and were increasingly embraced by the data mining community (Chapman, et al., 1999). These standards afforded the adoption of data mining by other groups, creating a reciprocal environment. As data mining tools, standards, and acceptable uses became more stable, the ability to employ data mining effectively increased, and as it did the tools, standards and expectations subsequently improved again. This phenomenon manifested itself in 2002, with Ng & Han's publication of the CLARANS methodology, which enabled the clustering of three dimensional spatial objects such as data cubes.

Data mining now could be found in highly sophisticated and intense quantitative research, including multidimensional and spatial research such as that found in Geographic Information Systems, Physics, and similar fields (Ester, Frommelt, Kriegel, & Sander, 2000; Han, Altman, Kumar, Mannila, & Pregibon, 2002). Scientists and social scientists today turn to data mining to address many of their research questions as a result of the progress which has been made. Often, this takes the form of simply applying the tools and standards which by the year 2006 have now arrived at a point of relative stability. At other times however, these new data mining consumers have enlisted the assistance of a growing pool of talented data mining experts (Jones, 2002) to create application-specific tools, algorithms or implementations which specifically address some research project (Han, Hu, & Cercone, 2003). Perhaps in no industry has this become more apparent than in health care (Harper, 2005). In Harper's study, data mining methodologies including association rules, logistic regression, clustering and decision trees were examined in the context of health care utilization and services. Previously

unrecognized patterns were found and strong recommendations followed for health care providers to take advantage of appropriate data mining tools in order to increase service levels, manage insurance premiums, and streamline processes in order to realize cost efficiencies. The high density of products and services provided by the health care industry demands a highly capable set of analysis tools which are now being delivered through data mining (Kyuseok, Srikant, & Agrawal, 2002).

Not to be outdone, marketing and sales operations have extended their own uses of data mining technologies in the past few years. Ultra-target marketing has become vogue, with association and correlation algorithms provided product suggestion services online, at the checkout stand, and even in print media (Ji, Liu, Sha, & Zhong, 2005). These systems, research has found, are increasingly inexpensive and simple to install, configure and maintain while providing a high return on investment and enhancement to customer satisfaction (Rygielski, Wang, & Yen, 2002). These techniques are having mixed results in the wireless communications market, where tying up limited or expensive cell phone and wireless PDA resources with promotional material is sometimes frowned upon by consumers (Saygin, & Ulusory, 2002).

While still a young discipline, data mining appears to be coming full circle now as its application is found in the very institutions which contributed to its genesis. Powerful computerized fraud detection and prevention engines, all built upon data mining algorithms and methodologies are today running non-stop, churning through hundreds of millions of data records generated through credit and debit card transactions and online purchases (Shi, Peng, Xu, & Tang, 2002). Financial institutions have long provided the catalyst for advancements in computerized information management and analysis, as they have sought effective mechanisms

for evaluating vast quantities of data for potentially fraudulent activity. Shi, et al. have found that not only is this being accomplished, but it is done in real-time—data is mined right as it is generated at the cash registers and online shopping cart checkouts. Even computer scientists themselves, who provided the knowledge of hardware, software and algorithmic foundations necessary for data mining to be realized now employ data mining in their own work.

Zimmermann, Weibgerber, Diehi, & Zeller (2005) cite their own success in using data mining to evaluate source code changes in order to guide future software programming decisions. Ying, Murphy, & Ng, (2004) extend this concept even further, by finding useful predictive data mining models in their evaluation of change history records. By understanding the software's history and evolution, Ying, et al. suggest that programmers can more effectively determine which future modifications will be most effective, well-received, and successful. This in turn will lead to greater consumer satisfaction, more acceptable results from the software itself, and lower overall software management costs.

Data mining has reached a point of usefulness to a broad constituency. This is due largely to the history of the discipline, which while not without its growing pains, has been a rapidly maturing one. The intentional and thoughtful recognition of the strengths and weaknesses of data mining, coupled with concerted efforts to organize, standardize, and simplify the processes involved, have led to a widely used and richly diverse user base. This foundation provides a bright future for the field of data mining.

### Learning Styles Research

In this study, a key component of the outcome is to understand student performance in light of their preferred, self-described learning styles. By understanding how students prefer to learn, and how they believe they learn most effectively, we can begin to understand which

pedagogical approaches are most appropriate for effective delivery of education to a target audience (Grashna, 1972). Here, this would entail the delivery of data mining education to the intended audience of non-Computer Science undergraduate students.

### *Definitions, History and Development*

The definition of learning styles has evolved over the past 30 years. In the late 1970s, the term *cognitive style* (Cohen, 1969) was perhaps a more common descriptor of the process of perceiving and processing information in order to understand, remember, and solve problems (Claxton & Ralston, 1978; Kirby, 1979). The terms *cognitive style* and *learning style* were used somewhat interchangeably during the early part of the 1970s, however much of that changed with the publication of David Kolb's *Learning Styles Inventory (LSI)* (1976), which was further refined and documented over approximately the next ten year (Kolb, 1985). Kolb's inventory provided a research-driven instrument with which to assess and categorize learners' preferences and styles of learning.

The use of the term *learning style*, according to Kirby (1979), is particularly appropriate for this study because it is consistent with the terminology used in Kolb's instrument, and it emerged from previous definitions of *cognitive styles* as a way to describe the process of *matching instructional materials to the needs of individual learners*, which is what we seek to do here. Reichmann (1978) would seem to support this position, through his claim that learning styles are defined by the learner's beliefs, attitudes and knowledge in the context of the material to be learned. It is through this literature that we arrive at the operational definition of learning styles to be used in this dissertation, as defined in chapter one.

Learning styles have emerged from and developed by and through the influence of different epistemologies. In the early days of learning research, these influences originated

primarily from the perspectives of behaviorism and cognitivism. Reichmann's perspective on learning styles originated primarily in behaviorist theory, with a strong focus on repetition, conditioning and response. Others however focused patterns of perception, recognition and organization, influenced more heavily by cognitive theory (Scarpaci & Fradd, 1985). Still others took a blended approach from these epistemologies. Keefe (1979) offered a concept of learning styles cognizant of both behavioral *and* cognitive beliefs when he discussed learning styles as both "physiological and psychological factors" which influence perception, interaction and response within the learner's environment.

During the 1980s and early 1990s additional refinement of the study of learning styles occurred. In 1980, Canfield published the *Learning Styles Inventory Manual*, which serves as an example of early attempts to bring standardization and a relative degree of consistency to the terminology and methods used in learning styles research. Kolb's 1985 update to his LSI aided in this effort, however Claxton and Murrell found that while certain level of agreement had been achieved by 1987, a single and concise definition for learning styles was unlikely without additional research and more well-defined theoretical foundation from which to work (Swanson, 1995).

The emergence of constructivist epistemologies shortly after the publication of Claxton and Murrell's work stands as one example of efforts to accomplish the desired theoretical foundation (Jonassen, 1990). This educational perspective embraced Keefe's earlier thoughts on a blend between behavioral and cognitive theories of knowledge and learning with the ultimate goal of creating instructional design models which attempt to incorporate and accommodate student learning styles (Duffy & Jonassen, 1992; Jonassen, 1994). This new focus on the interaction between the student's predispositions, knowledge and experience and the

methodologies with which we teach brought learning styles research to a new threshold, which ushered in a new wave of scholarly inquiry into learning styles (Cornwell & Manfreda, 1994).

The 1990s and 2000s have been characterized research in the statistical validity and reliability of various learning style instruments (Cockerton, Naz, & Sheppard, 2002; Koob & Funk, 2002). Allinson (1990) conducted some of the earliest research in the areas of validity and reliability, uncovering areas for potential improvement of several of the most common learning styles instruments. Two key studies followed which are specifically relevant in that they examined validity and reliability of Kolb's updated LSI which had been released in 1985. Loo's findings (1996) tested for and found improvements in both statistical validity and classification reliability in Kolb's 1985 LSI over the earlier 1976 version; while research from De Ciantis and Kirton (1996) sought to extend Kolb's categorization of individual learning styles into strata of level, style and process, thus refining the data obtained from the instrument. Efforts such as these, conducted over a period of about ten years with similar results, addressed major concerns about learning style assessment instruments and helped to solidify their adoption within the research community (Kayes, 2005). Validation of processes and instruments related to learning styles research ultimately helped paved the way for additional studies, as confidence in learning styles research has provided a natural segue to a connection with instructional models (Boyle, Duffy, & Dunleavy, 2003).

Acceptance of learning style instruments and a desire to match teaching methodologies with learning styles increased dramatically in the past decade. Educators wanted to know how the two interacted—a question researchers had raised many years earlier (Stensrud & Stensrud, 1983). Sayer and Studd (2006) recently published a study which illustrates this point, finding that even among the most homogenous group of learners, learning styles tend to be variable, however

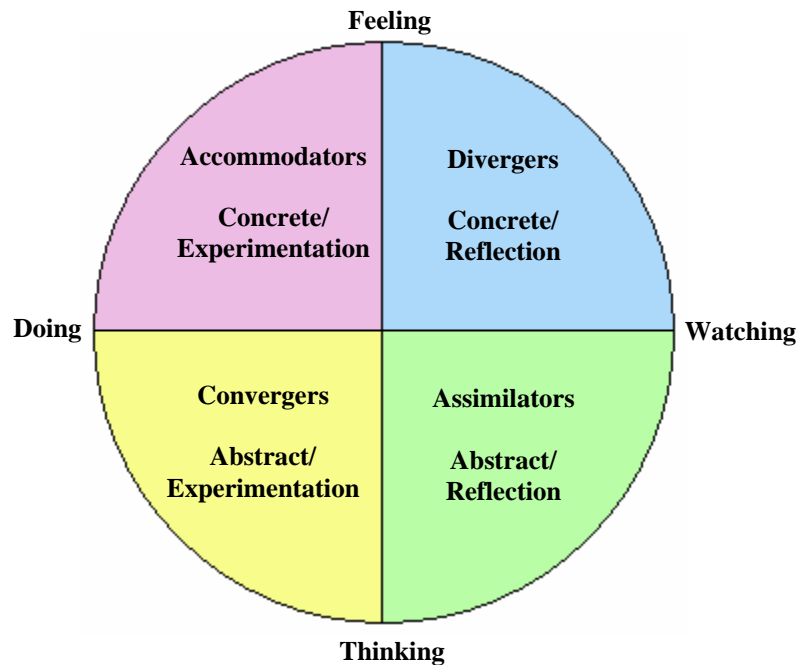


their findings indicate that this does not need to become an impediment to instructors identifying and addressing the learners' styles at some instructional level. Efforts at employing the *most suitable* teaching models for the students' learning styles yielded positive outcomes for their students. It is therefore important to understand how Kolb's LSI stratifies learners into learning style categories.

### *Kolb and Categorization*

At the heart of learning styles research is the categorization of each individual learner into a descriptive learning style which best matches that person's preferences and abilities to understanding, codify, recall and apply new knowledge (Mainemelis, Boyatzis, & Kolb, 2002). From the beginning of Kolb's work in this area, he had conceptualized four categories or styles of learning, and posited that every person could be categorized into one of these areas, although no learner would necessarily be defined discretely by their self-described category (Kolb, 1976; 1985). Rather, the learner would be placed somewhere along a continuum which most closely related their personal learning style to one of four categories: diverger, assimilator, converger or accommodator (see Figure 2.1).

Figure 2.1. Kolb's Learning Style Categories



Sharp (1996) summarized the definitions for each of Kolb's categories:

*Divergers:* Individuals within this category rely heavily upon their senses to facilitate learning. They seek to understand a problem or situation from various perspectives and value information which is personally applicable. Learning for these individuals is most effective in an environment where they can observe and employ other sensory mechanisms in order to acquire and reflect upon new knowledge. In Kolb's terminology, people in this category prefer concrete concepts and reflection.

*Assimilators:* Learners who fit this category are rational thinkers. They are capable of more abstract thought and seek to couch new knowledge in the context of precedent, procedure and theory. Once a new concept is attained, assimilators are generally capable

of integrating the new information into a larger contextual view. In Kolb's terminology, people in this category prefer abstract concepts and reflection.

*Convergers:* Those who learn through convergence are characterized by a tendency toward experiential and experimental learning. These individuals actively seek a "correct" solution, often by testing abstract ideas in order to see rather than to be told about the outcome. In Kolb's terminology, people in this category prefer abstract concepts and experimentation.

*Accommodators:* Individuals who learn through accommodation acquire new knowledge through specific examples and illustration, and then adapt their learning through conceptual application or proof-of concept activities. Accommodators adapt more easily than those in other categories to new ideas or ways of applying knowledge. In Kolb's terminology, people in this category prefer concrete concepts and experimentation.

This dissertation does not seek to suggest beforehand which of Kolb's learning style groups, if any, will realize the highest assessment outcomes from the data mining instruction to be provided. Instead, the null hypothesis is that there will be no statistically significant difference among the learning style categories. If in fact we do reject this null hypothesis, a post hoc evaluation of the ANOVA results will reveal which of the learning styles differ significantly from the others during the acquisition of new knowledge in the field of data mining.

### *Online Learning*

Because the learning in this study will take place online, it is important to contextualize the online learning environment in light of current literature. An exhaustive review of online learning literature would not serve any specific purpose in support of the current study, however it is important to recognize key components which may influence or otherwise impact the data or

findings in this study, particularly as these may relate to impacts on learning styles in the online learning environment.

In their study, Mupinga, Nora, and Yaw (2006) found that a lack of technical readiness or unfamiliarity with an online learning environment might be confused with a learning style which is not highly compatible with online learning. It is therefore important if one is to consider accommodating learning styles during the design of online instructional materials that students be prepared to overcome any technical limitations they may have. Failure to engage in such preparatory activity may result in a masking of actual outcomes—in other words, what may appear to be a poor student outcome due to low compatibility between the learner's preferred style and the instructional delivery model may actually have been caused by the student's inability to learn the material in the online environment. The implications are that students need to be as prepared as possible, perhaps through a brief tutorial, in order to learn the material through the online interface with a minimum of interference or confusion caused by the online environment itself. The risk of such confounding events is low in this study due to the demographic nature of the study participants. This risk will be further controlled for by collecting a small amount of data related to each participant's level of comfort and familiarity with online technology. This method of control, as illustrated in Frederico (2000) in which Kolb's inventory was specifically used, is useful in identifying any observation in the resultant data set of a study which has both low outcome scores and low levels of comfort or familiarity with online learning environments. Participant scores found to be unduly influenced by technological limitations are then not included in calculations and analyses of the study.

Other potentially confounding influences do exist in the online teaching and learning environment, however the risk of such issues affecting this study are also low. Learner

motivation, for example, is a constant concern in the online education environment (Fritz, Speth, & Barbuto, 2004), however, since the data mining instruction for this dissertation will occur in a single sitting it is unlikely that student participants will be affected by a lack of motivation. Other traditional barriers including unreliable technology or personal distractions are also limited due to the short-term and focused nature of the instructional/learning events required for this study (Zirkle, 2001). Through the intentional consideration of these typical known factors in online learning, we will be able to create an experience which will yield useful data for this study, capturing true learning outcomes which can be compared to each participant's learning style (Baldwin & Sabry, 2003; Miller, 2005).

#### *Summary of Learning Styles Literature Review*

The study of learning styles emerged primarily from cognitive studies in the 1960s, with some degree of influence from behavioral theory as well. This cited research demonstrates that Kolb's work in providing the Learning Styles Inventory contributed a level of consistency and reliability to the field which helped to legitimize and substantiate the research coming out of it. Efforts to validate Kolb's instrument, particularly the 1985 updated inventory, have resulted in greater degrees of confidence in learning styles research as well—Kolb himself has conducted some of this very research, in an effort to further support and refine previous findings and theories (Mainemelis, Boyatzis, & Kolb, 2002). As mentioned in Sayer and Studd (2006), research in the field today has shifted more toward the interaction between students' learning styles and models of teaching embraced by the instructor. In this study, the teaching and learning will take place online, and although this situation does come with an added dimension of potentially confounding variables, these are either not particularly relevant or will be mitigated specifically during the execution of the methodology of this study.

## Models of Teaching

The second null hypothesis of this dissertation suggests that the model of teaching within which data mining instruction is delivered to the non-Computer Science target audience will have no effect on student performance. Educational psychologists have long debated and researched on the topic of models of teaching in search of the most effective ways to teach specific content or to deliver instruction in the context of learners' needs. It is important to examine the role of teaching approaches in teaching and learning, discuss the specific models that have been selected for this study and why, and address the dynamics introduced by the online teaching environment.

### *Review of Models of Teaching*

Like learning style theory and research, models of teaching emerged in the 1960s, also spurred by research related to the theories of behaviorism and cognitivism. Wilson and Cole (1996) explain that behavioral research really provided the early impetus for teaching model research, beginning in about 1960 and continuing through the mid-1970s. Cognitive theorists began to participate very actively in teaching model research by 1975, and this activity has continued through to today, however the aforementioned shift toward social constructivism (Jonassen, 1990, 1994) also impacted teaching model research beginning around 1990. In each of these three phases defined by Wilson and Cole, the relationship between instructional design and instructional psychology underwent notable shifts:

- 1960-1975: Behavioral phase of teaching model research
  - Instructional Design / Instructional Psychology relationship: The two are closely aligned, models of teaching begin to emerge, theories on linkage between

approaches to teaching and style of learning are suggested (Gagné, 1968; Merrill & Tennyson, 1977).

- 1975-1989: Cognitive psychology phase of teaching model research
  - Active teaching model development: Instructional Design diverges from Instructional Psychology as designers focus on model development and teaching/learning interaction while psychologists gravitate more toward the “cognitive mainstream” (Reigeluth, 1987; Resnick, 1981).
- 1989-present: Knowledge construction phase of teaching model research
  - Reengagement of Instructional Design and Instructional Psychology: As teaching model research has matured and also entered the mainstream, and as cognitive psychologists have followed the mainstream toward constructivism, the two sides which provided the genesis of teaching model research have more closely realigned with one another and renewed discourse (Jonassen, Bessner, & Yacci, 1993).

As with any research pursuit, something must provide a catalyst for continued investigation, development, and experimentation. No exception, research on models of teaching has been primarily driven by the desire to understand teaching effectiveness and the degree of successful learning achieved by the learner (Joyce, 1980; Lieberman, 1982). This has been a common theme through the three phases of teaching model research outlined above.

During the behavioral phase, much attention was paid to the relationship between instructional intervention and subsequent student performance (Hunt, 1974). The general theory which dominated the field during this time suggested that improved teaching methods could be equated to increase student performance (Skinner, 1986). While much research provided

evidence to support this theory, many, particularly cognitive psychologists remains dissatisfied with explanations which appear to overly credit the teaching model with successful learning (Collins & Stevens, 1983). The resultant effect produced additional models of teaching which sought deliver effective teaching through and understanding and addressing of the physiological and psychological processing of acquiring, understanding, remembering and applying information (Dillon, 1998). With the development and acceptance of social constructivist epistemologies in the 1990s, a diverse array of approaches to teaching developed and began to spread and form into teaching model “families”, which often overlapped in some areas while addressing very specific areas related to content, audience or environment (Palmer, 2005). Allen (1996) and Joyce, Weil and Calhoun (2000) contributed significantly to the codification, organization and publication of these teaching families, which are summarized in Table 2.1.

Table 2.1

*Teaching Model Families*

Model Family	Objective(s) of the Family	Sample of Models or Approaches
Behavioral	Modify behavior of the learner through deliberate and programmatic manipulations of specific knowledge, skills and/or aptitudes.	<ul style="list-style-type: none"> <li>• Direct Instruction</li> <li>• Simulation</li> <li>• Drills</li> <li>• Programmed Exercise</li> <li>• Conditioning</li> </ul>
Cognitive	Develop within the learner logical	<ul style="list-style-type: none"> <li>• Concept Attainment</li> </ul>



	and problem solving abilities and processes for critical thinking.	<ul style="list-style-type: none"> <li>• Memory Models</li> <li>• Case-based Reasoning</li> <li>• Problem Solving</li> <li>• Apprenticeship</li> </ul>
Social Constructivist	Create building blocks of previously acquired and stored knowledge with which to create a framework for the acquisition, contextualization, and application of new knowledge or skills.	<ul style="list-style-type: none"> <li>• Brainstorming</li> <li>• Cooperative Learning</li> <li>• Socratic Method/Inquiry</li> <li>• Group Discussion and Response</li> <li>• Nondirective Project</li> </ul>

---

The formalization and stabilization of teaching models paved the way for more pointed research, such as this dissertation, into the relationship between mode of instruction and style of learning. Research continues to support the need for continued inquiry into the teaching model/learning style interaction, particularly within the context of a specific content (Searson & Dunn, 2001; Westman, 1993).

### *Mapping Models of Teaching to Learning Styles*

As early as 1979 (Ellis), the need for integration of models of teaching and learning styles was recognize as an emerging problem. As researchers began to understand more about both aspects of the instructional environment, additional efforts to merge and correlate the two accompanied this new understanding. Ellis suggested a solution which departed from the norm of attempting to match students with teachers who most closely aligned with the students'

learning styles. Instead, Ellis' approach was to arm teachers by providing in service instruction in a variety of teaching models, allowing the teacher to adjust and adapt teaching approaches to students and to content to be delivered.

In spite of early efforts such as Ellis', the alignment of teaching methodologies with learning styles remains a question with many answers, or at least ideas. In 1997, Thompson posed the question of who should be adapting to whom. Should the onus of matching the instructional experience to the learners' preferences rest solely on the teacher? Or is it reasonable to expect that given the broad range of students in a given class will have an equally broad range of learning styles among them? If the latter is the case, and according to Thompson it often is, then the traditional belief that the instructor ought to adapt their teaching methodologies to match learning styles is flawed, because there is no way to adapt teaching to work equally well for multiple learning types simultaneously. Therefore, Thompson concludes that it is reasonable to expect learners to adapt their preferences to match with the teaching model used, and that the teaching model should be carefully selected to match both the material to be presented and the *most prevalent* learning styles among the collective group of learners.

Ample scholarly support exists for Thompson's recommendations, particularly in relationship to matching teaching models with content. As previously cited, Westman (1993) offers such support by claiming that learners often have different learning styles which vary dependent on the subject matter being taught. For example, in learning sculpture, a learner may be classified as an accommodator, preferring to learning in a highly hand-on and experiential way (e.g. hands on clay), while that same person may prefer to learn computer programming as an assimilator, processing the concepts of recursive loops or if-then-else logic in an abstract way before sitting down at a computer keyboard to try writing such statements.

Gershoff (2005) agrees that learning styles are chosen within the context of the material being taught, but suggests that through instructional decisions which allow students to self-select learning activities or work collaboratively to complement one another's individual learning styles instructors can successfully "help students of all learning styles succeed". This study is presented specifically from the perspective of Political Science instruction. Similar findings from the fields of Business and Computers support Gershoff (Loo, 2002; Ross, Drysdale & Schulz, 2001), however other published information suggests that instructional decision-making can only go so far in mitigating the disparity of learning styles among students (Enkenberg, 2001). In spite of the instructor's best efforts to ensure a high degree of compatibility between teaching approach and the learning styles found among class participants, no instructional design will completely close the gap. Teachers and students can however become active participants in addressing the teaching model/learning style dilemma by becoming conscious of both (Chittleborough, Treagust, Mamiala, & Mocerino, 2005), and then mapping the instructional content to the teaching and learning approaches being used in a given situation.

### *Selecting Models for this Study*

Han & Kamber (2001) list Classification and Prediction as two fundamental areas of data mining which all data mining students should learn. This study will use a learning unit from each of these two data mining areas as the instructional content to be delivered to the study participants. The selection of Direct Instruction and Concept Attainment as the experimental teaching models in this dissertation is intentional in light of the nature of this instructional content.

### *Direct Instruction*

The term “Direct Instruction” was introduced to describe a behavioral model of teaching wherein the teacher leads the learning through delivery of specific instructions and practice, and then follows up by assessing the students’ ability practice independently and by providing immediate corrective feedback when necessary (Joyce, Weil & Calhoun, 2000; Rosenshine, 1976). This model has proven most effective in training learners in specific tasks which are highly structured or otherwise stratified into a predictable number and/or sequence of steps (Adams & Engelmann, 1996). Stein, Carnine and Dixon (1998), break Direct Instruction down into five phases or principles:

1. identify "big ideas" to organize content;
2. teach explicit, generalizable strategies;
3. scaffold instruction;
4. integrate skills and concepts;
5. and provide adequate review.

Stein, et al.’s approach to Direct Instruction is supported by years of instructional design research, which in turn supports the choice for this teaching model as one of two in this dissertation.

Within the data mining area of classification there are a number of data mining models, algorithms and approaches. One of the most widely used is Association Rule mining (Ji, Liu, Sha & Zhong, 2005). Association Rule mining is the process by which correlations between items are examined in order to determine the strength of their relationship. The strength of relationship extends beyond a simple correlation however to include frequencies and proximities of items in a data array, which are algorithmically evaluated in order to find the items which are

most alike. The results show products which may be marketed together, behaviors which when observed together indicate a theme, or processes which are closely aligned and may be candidates for combination or streamlining. The pragmatic process by which data are prepared, organized and computationally evaluated for strength of relationship makes this data mining learning unit a prime candidate for a Direct Instruction approach.

### *Concept Attainment*

Concept Attainment is a cognitive/constructivist model of teaching in which learners are introduced to a topic and taught basic rules for differentiating exemplars of that topic from non-exemplars (Allen, 1996). Student must use their own cognitive maps, often relying upon previous learning in order make successful decisions about which items in a set belong, and which do not. This teaching approach has been found effective across all age groups (Joyce, Weil & Calhoun, 2000), although adult learners have shown the ability to attain higher levels of sophistication in their learning through this model (Lewis, 1980). Association Rules within data mining are logically constructed based upon the relative frequency with which items are found together in the data set. In order to identify pairs or groups of items as “rules”, each record in the data set must be examined in comparison with every other record. This step-by-step process of identifying representative and non-representative items (or matches) is strongly compatible with the Concept Attainment teaching model.

### *Online Teaching*

Although the main objective of this dissertation is to search for interactions between learning styles and teaching models, the instructional experience for the study participants will take place online. Many of the concerns surrounding the online component of this study have

been addressed previously. In the context of online *teaching* however, it is worth noting that while increased online literacy and experience of the learner have contributed to higher student performance, significant differences have not generally been observed related to the use of Direct Instruction or Concept Attainment when teaching online (Martindale & Ahern, 2001; Sims, Dobbs & Hand, 2002). Dowling, Godfrey and Gyles (2003) have suggested that the use of online teaching tools can improve student outcomes, however in their study the online teaching was used as a supplement to traditional face-to-face teaching. Because this study will deliver and assess participants entirely online, based upon Sims, et al. and Martindale and Ahern, it is believed that the teaching models will not be significantly influenced by the fact that they are being used online.

#### *Summary of Teaching Model Literature Review*

Teaching model research has been strongly influenced by instructional psychology and instructional design research for more than the past 30 years. Direct Instruction and Concept Attainment have emerged as two widely accepted models which have been identified as excellent candidates for the instructional units to be used in this study. The use of these two models is central to the outcome of this dissertation, as we seek to understand whether or not there is any interaction among the two teaching models and Kolb's four learning styles within the context of introductory data mining instruction.

#### The Pedagogy of Data Mining

Urbancic, Skrjanc, and Flach (2002) conducted an extensive web-based analysis of data mining and decision support education, however this study primarily focus on course offerings and target audiences rather than on pedagogical issues such as learning styles or teaching

models. Other studies have focused on the importance or relevance of teaching a certain data mining topic, such as ethics (Morse & Morse, 2002), but without focusing specifically on the models of teaching or learning. Several other studies have employed data mining methodologies as a means of assessment or evaluation of teaching effectiveness, though these range across a broad spectrum of curricula including Special Education, Biology and Engineering (Gwo-Jen, 2005; Tsantis, & Castellani, 2001). Some studies within this vein have even employed data mining as an evaluative tool in online teaching and learning (Chen, Ou, Liu, & Liu, 2001; Dringus & Ellis, 2005; Shen, Hart, Yang, Yang, & Zhexue 2003). Cohen (2003) offers perhaps the most useful insight into the relationship between data mining and the improvement of the teaching/learning interaction, finding that data warehousing and data mining can be used to reveal patterns of relationship between teachers and student performance scores, however this is again largely in the context of program evaluation across myriad fields of study. This dissertation will complement and expand upon all previous research, thereby adding to the body of knowledge within the fields of data mining education, teaching models and learning styles, with potential to also address interesting findings (if any) related to the online teaching and learning environment.

### Summary of Literature Reviewed

We seek to investigate the interaction of Kolb's four learning styles within specific data mining instruction delivered through Direct Instruction or Concept Attainment models of teaching. The review of the history, appropriate uses and previous research findings in the areas of learning styles and models of teaching provides justification and necessary practical framework in support of this study. In light of existing literature on the subject of *data mining education*, this study is further justified, given its potential contribution to the field.

### CHAPTER III: METHOD AND PROCEDURE

As discussed in Chapter 1, we hypothesized the following for this study:

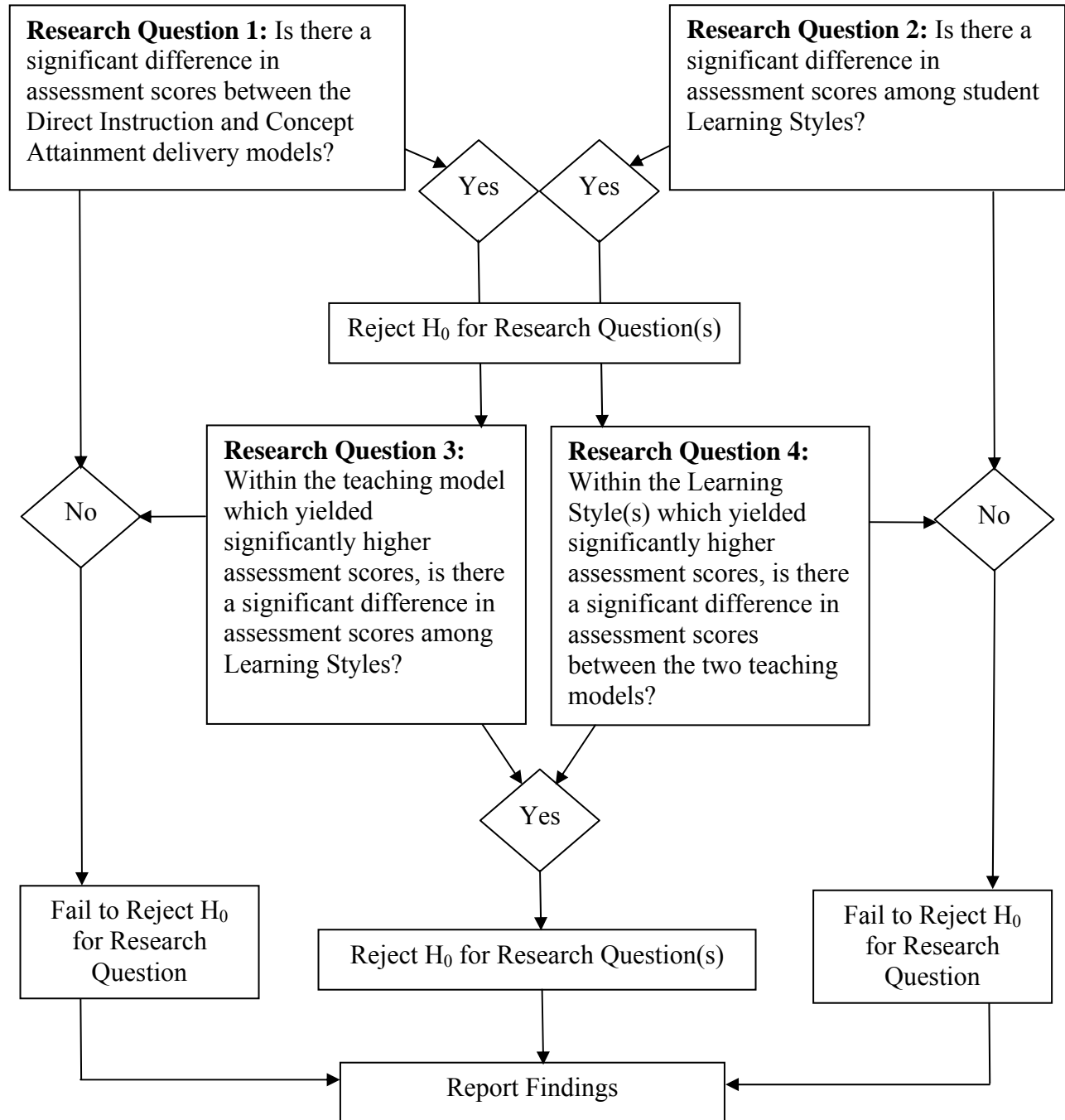
1.  $H_0$ : There is no significant difference in assessment scores when either Direct Instruction or Concept Attainment is used as the teaching model in undergraduate, non-Computer Science data mining instruction.
  - a.  $H_a$ : There is a significant difference in assessment scores when Direct Instruction or Concept Attainment is used as the teaching model in undergraduate, non-Computer Science data mining instruction.
2.  $H_0$ : There is no significant difference in assessment scores among various learning styles in undergraduate, non-Computer Science data mining instruction.
  - a.  $H_a$ : There is a significant difference in assessment scores for students categorized in certain learning styles in undergraduate, non-Computer Science data mining instruction.
3.  $H_0$ : If in Research Question #1 above,  $H_a$  is true, there is no significant difference in assessment scores associated with students' learning styles within the teaching model which yielded the significantly higher scores.
  - a.  $H_a$ : If in Research Question #1 above,  $H_a$  is true, there is a significant difference in assessment scores associated with students' learning styles within the teaching model which yielded the significantly higher scores.
4.  $H_0$ : If in Research Question #2 above,  $H_a$  is true, there is no significant difference in assessment scores between the two teaching models which can be associated with students' learning styles.



- a.  $H_a$ : If in Research Question #2 above,  $H_a$  is true, there is a significant difference in assessment scores between the two teaching models which can be associated with students' learning styles.

These hypotheses are further illustrated in Figure 3.1.

Figure 3.1. Dissertation Hypothesis Map

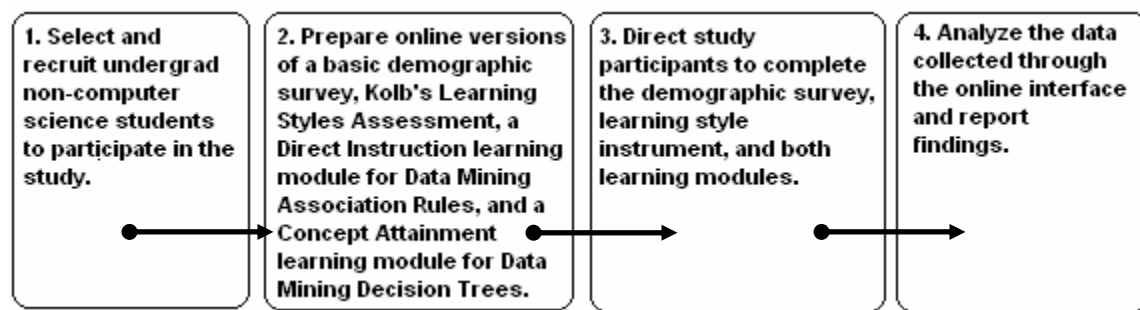


In order to address each of the four research questions, the design for this study includes an organizational model, a description of participants, the development of data collection instruments, a description of data collection methodology, and finally, an outline of data analysis approaches.

### Organizational Model

Figure 3.2 illustrates the organizational model which enabled timely and accurate preparation of data for this study.

*Figure 3.2.* Organizational Model



All data for the study was collected through the online interface, which is further documented in the Instruments section below (code provided in Appendix D). The data was stored on a secure server and evaluated strictly within that environment. Because human participants were involved, this study was subject to approval as an exempt study by the Institutional Review Board (IRB) of the schools where the research took place. IRB approval forms required are included as Appendix A of this dissertation.

### Participants

In the operational definitions portion of this document, the study participants are defined as undergraduate college students not majoring in Computer Science at a traditional, residential

teaching college. These participants engaged in the study on an entirely volunteer basis and were not personally identifiable through the data collected. They were informed of such at the beginning of the online session. Participants were recruited with a resultant sample of 86 participants completing all of the online portions of this study. No incentive, monetary or otherwise, was offered to entice participation. Demographic data is reported in order to afford readers of this dissertation a more informed position as to the inferential nature of this study to other groups or populations.

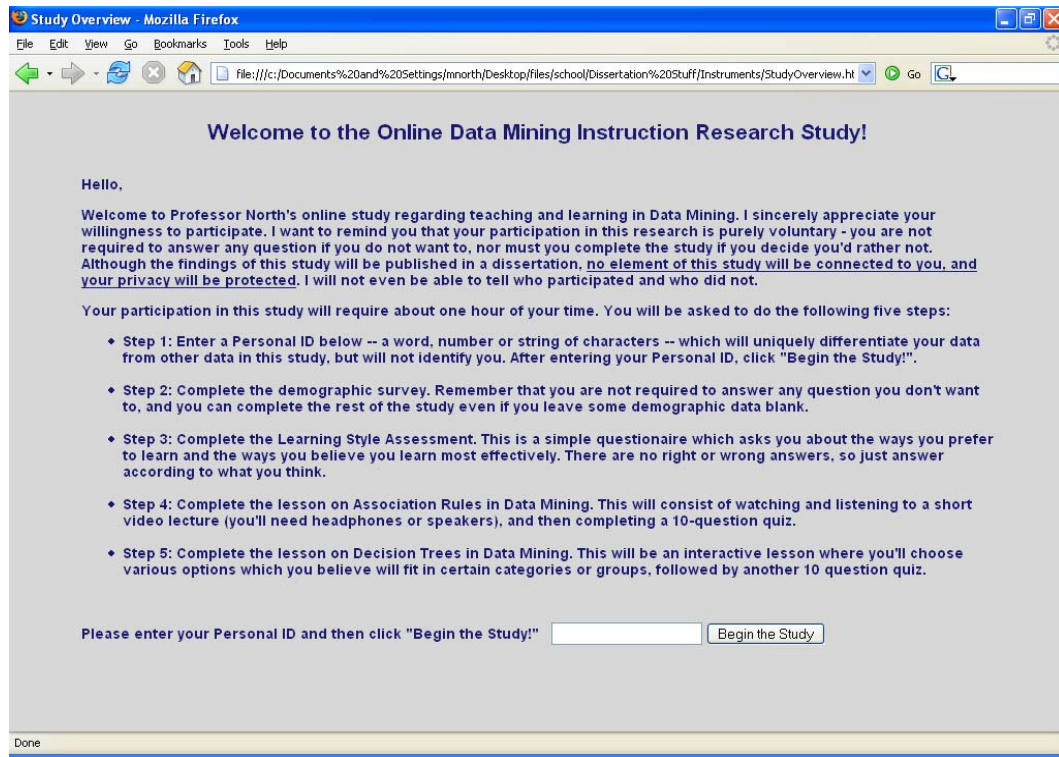
### Instruments

Study participants were given the Web site address to begin the study at the time they agreed to participate. They were encouraged to set aside at least 30 minutes to complete the process. Upon accessing the Web site, the students worked through four instruments in the following order: Study Overview, Demographics, Learning Styles, and Rules of Association. For the final instrument, study participants were randomly selected into either a direct instruction or concept attainment instructional environment, and after completion of the learning activities, all completed the quiz assessing their understanding of Rules of Association.

#### *Study Overview*

Figure 3.3 depicts the first instrument which study participants reviewed. Its purpose was to remind them of the voluntary nature of their participation and of the protection of their privacy, and to uniquely identify them by a user-defined character string. The Participant ID was written to the database from this instrument.

Figure 3.3. Study Overview and Participant ID Registration



### Demographics

Figure 3.4 depicts the demographic instrument which study participants completed. Its purpose is to allow for stratification of the data in the study along potentially relevant demographic lines. Demographics such as race, religion, disability, etc. not considered within the scope of this study were intentionally excluded. All fields shown in Figure 3.4 were written to the database from this instrument.

Figure 3.4. Demographic Data Collection Instrument

The screenshot shows a web browser window titled "Study Overview - Mozilla Firefox". The address bar displays a file path: `file:///c:/Documents%20and%20Settings/mnorth/Desktop/files/school/Dissertation%20Stuff/Instruments/Demographics.htm`. The main content area is titled "Demographic Data". On the left, a green box labeled "Progress..." contains a list of five steps: 1. Register Personal ID (Done), 2. Demographic Data (In Process), 3. Learning Styles (Pending), 4. Register Personal ID (Pending), and 5. Register Personal ID (Pending). The main form area contains the following fields: Age (17), Gender (Female), Academic Year (Freshman), Academic Major (Accounting), Internet Proficiency (Low), and Highest level college math completed (100). A "Save Demographic Data & Continue" button is located at the bottom right of the form.

Field	Value
Age	17
Gender	Female
Academic Year	Freshman
Academic Major	Accounting
Internet Proficiency	Low
Highest level college math completed	100

### *Learning Styles*

Figure 3.5 depicts Kolb's Styles Inventory, which study participants completed. Its purpose is to collect the self-described learning styles for each study participant. All fields in the inventory were written to the database from this instrument. The full instrument is included as Appendix B of this dissertation.

Figure 3.5. Kolb's Learning Styles Inventory

**Learning Style Inventory**

**Progress...**

1. Register Personal ID (Done)
2. Demographic Data (Done)
3. Learning Styles (In Process)
4. Rules of Association (Pending)
5. Decision Trees (Pending)

Please answer the following questions as thoughtfully as you can and then click "Find My Learning Style".

I prefer

- ☐ hands-on learning experiences.
- ☐ learning through thinking and reasoning.

I learn well by

- ☐ practical experience.
- ☐ applying theories to hypothetical situations.

I tend to

- ☐ jump right in and do something new.
- ☐ think about possible outcomes before trying something new.

I learn more effectively from

- ☐ my peers.
- ☐ my teachers.

I learn best through

- ☐ active involvement in projects.
- ☐ observation.

I like learning through

- ☐ simulations.
- ☐ lectures.

I tend to

- ☐ rely on feelings when making decisions.
- ☐ rely on logic when making decisions.

Done

### *Rules of Association via Direct Instruction*

Figure 3.6 depicts the introduction to the Direct Instruction learning module completed by half of the study participants. This learning module used Direct Instruction to prepare students for the ten question quiz on how to build and interpret the results of an Association Rule data mining model. The instruction was delivered as an audio/visual lecture and demonstration embedded in a Web page. Participants watched and listened as the concept of Association Rules was explained, and then as the process for building and interpreting an Association Rule model was demonstrated in a data mining software package called AlphaMiner. Figure 3.7 depicts the quiz which assessed the participants' learning of Association Rules in Data Mining. The full quiz is included as Appendix C of this dissertation.

Figure 3.6. Direct Instruction Learning Module for Association Rules in Data Mining

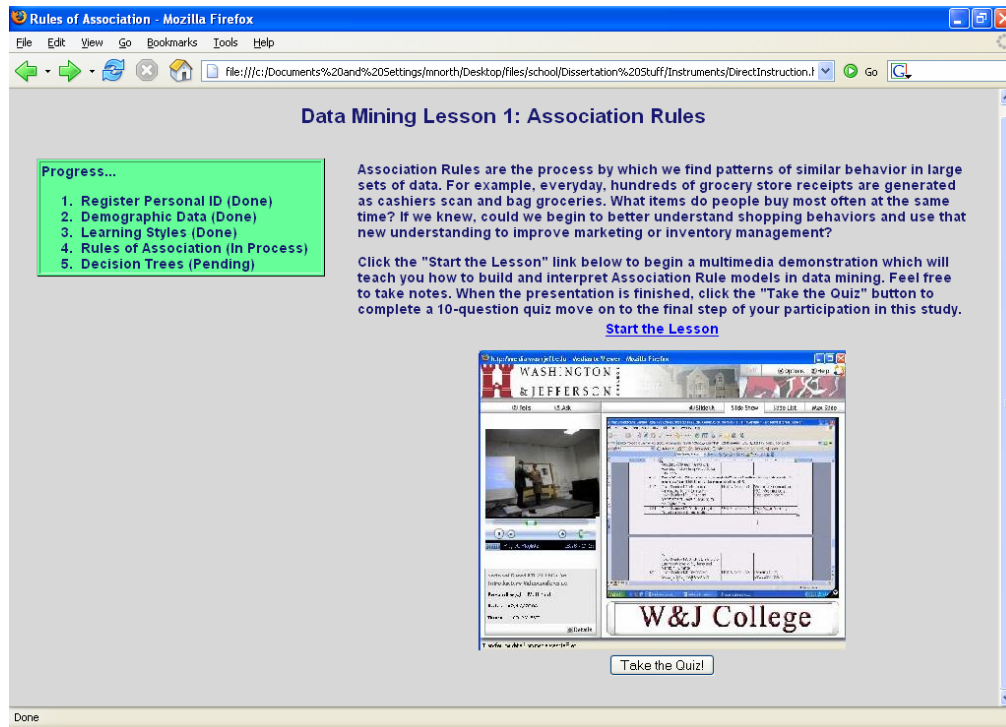
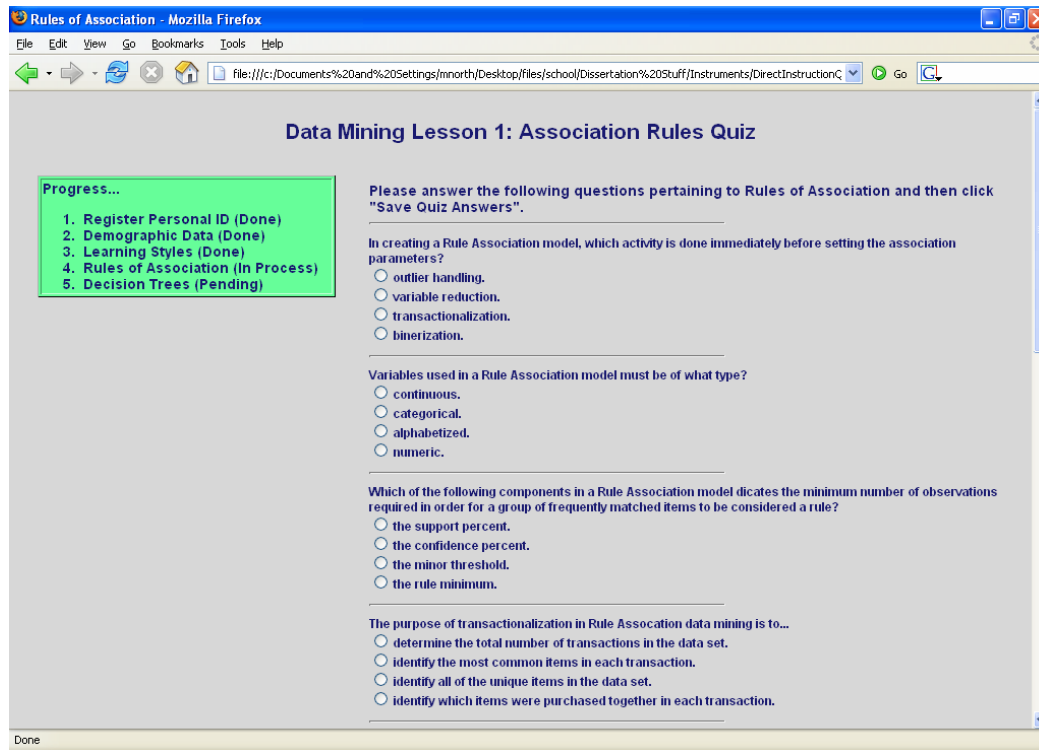




Figure 3.7. Assessment Instrument (Quiz) for the Association Rules Learning Module



### *Rules of Association via Concept Attainment*

Students were randomly be assigned to receive Association Rules instruction via either Direct Instruction of Concept Attainment. Figure 3.8 depicts the learning activity which uses the Concept Attainment model to prepare students for the quiz. The interactive concept attainment activity was embedded in a Web page and built upon Bruner's original cognitive model (1950):

1. Introduce the process to the students
2. Present the examples and list the attributes
3. Develop a concept definition
4. Give additional examples for practice
5. Evaluate

For this lesson, students were introduced to the concept of Association Rules by evaluating shopping cart products. Participants then built association rules by determining which shopping cart items were exemplars of frequent pairs, or in other words, rules.

*Figure 3.8.* Concept Attainment Learning Module for Association Rules in Data Mining



### Data Collection

All data for this study were collected in real-time as the study participants completed the online learning modules. The Web pages recorded the data as outlined in Figures 3.2-3.8 and stored this data in a database. No other data beyond that which has been described above was collected or used in this study. All data collected is included in Appendix E.

### Data Analysis

Once collected, the data for this study were analyzed in order to address the four hypotheses of this dissertation. Statistical tests were performed to determine the existence of significant differences between or among groups of learning styles or teaching models. Data

mining methods are also employed in exploratory data analysis in order to find the existence and strength of any relationships which may exist in the data.

### *Demographic Analysis*

The analysis of the demographic data is largely descriptive in nature in order to characterize the profile of the student participants in this study. Summary statistics including counts and measures of central tendency and dispersion are included where relevant in order to assist the reader in understanding the larger population to which the results of this study might later be inferred.

### *Analysis of Models of Teaching*

The first hypothesis in this study is designed to understand whether or not students will perform better based upon the model of teaching with which they were instructed. Since two specific models are being examined, a two-tailed *t*-test will reveal any significant difference between the two groups of quiz scores, as well as indicate which of the two methods resulted in higher scores. The outcome of the *t*-test allows us to respond directly to the first research question, and forms a justification for whether or not the third research question needs to be addressed as well. Following established educational and social science research norms, the alpha level ( $\alpha$ ) is set at .05 for all statistical tests in this study (Freund & Simon, 1997).

### *Learning Style Analysis*

Given that there are four learning styles in Kolb's inventory, an analysis of variance (ANOVA) was required in order to identify whether there is a significant difference between learning style groups in this study. The assessment scores from the two quizzes are compared across all four learning style groups to determine whether or not a significant difference exists

among them. A result showing significance merits post hoc comparisons in order to determine which of the groups yielded significantly higher quiz scores, which will address our second research question and will provide a foundation for addressing the fourth research question, if necessary. If a significant difference between groups were found during the ANOVA exercise regarding research question two, Scheffe's post hoc methodology would be employed in order to identify which groups scored higher. Because it will be impossible to ensure beforehand that all learning style groups will be represented equally in terms of sample size, Scheffe's methodology is therefore most appropriate for this particular study (Jones, 2006).

### *Level Two Hypotheses*

Research questions three and four would only need to be addressed if research questions one or two had yielded statistically significant results. If the *t*-test or ANOVA for research questions one and two, respectively, yielded statistically significant results, then the reverse application of these two statistical tests would have been applied in order to answer research questions three and four. Since no significant differences were found however, no additional statistical tests for these second level hypotheses were performed.

In addition to the statistical measures, as a complement to the findings, an Association Rule exercise and a Decision Tree exploration were conducted in order to determine whether or not there is any strength of association between the scores in each of the two models of teaching and each student's learning style. This exploratory data analysis is reported in the findings section in order to illustrate the lack of association between the scores and learning styles and was conducted using the standard Association Rule and Decision Tree algorithms.

### *Summary of Research Methodology*

The data for this dissertation was collected through an online learning system which was available only to student participants. Students who elected to participate in this study registered an anonymous and unique personal identifier, then provided a limited set of demographic information and completed Kolb's Learning Style Inventory. They then received a data mining lesson via the online interface through either Direct Instruction or Concept Attainment. The lesson was followed by a 10 question quiz as illustrated in appendix C.

The data collected were evaluated using three measures. *T*-tests are used to find statistically significant differences when teaching model data is in question, since only two groups exist in this area of interest. ANOVA was employed whenever learning styles were evaluated, and any significant difference found in the learning style data would be further examined using Scheffe's post hoc comparison procedure. Finally, the data were prepared and mined using Association Rules and Decision Trees to reveal the existence and strength of associations between quiz scores in each of the two teaching models and the learning styles of each participant.

## CHAPTER IV: RESULTS

The primary objective of this dissertation is two-fold: first, to determine whether or not one of the two target methods of teaching would result in significantly higher quiz scores in the introductory data mining lesson on Association Rules; and second, to determine whether or not any of Kolb's four learning styles would yield significantly different quiz scores among study participants. In the case that one or both of these questions is answered in the affirmative, the secondary objectives of this study are to determine if there are significant differences in quiz scores within the more successful model of teaching or learning style. This chapter outlines the findings of this study which address the four hypotheses as stated at the beginning of the previous chapter.

### Outcomes of Data Analysis

#### *Demographic Data and Discussion of Participant Profiles*

Students were recruited to participate in this study during the summer of 2006. The college at which this study was conducted does not have a Computer Science department, and thus any students who were enrolled and willing to participate were invited to join the study. Students were provided with a Web address on a college server and allowed to log on and complete the assessment at their convenience during a two-week window. In all, 106 students were invited to engage in the study, and 86 ( $n=86$ ) complete participant records were collected during this data collection period. As a lecturer at the college where this study was conducted, I had taught many of the students who were invited to participate. The participants were asked in person if they would help with the study, and if they agreed, directions to the online learning materials were then given to each participant. Although no enticement or compensation was

offered to these participants, this represents an 81.1% return rate which was higher than anticipated. Appendix D contains the code used to create the online data collection and instructional interfaces used by these study participants, while Appendix E provides the master data set which was collected through these instruments. In addition to collecting this data, the Web server logged the amount of time each participant took to complete the four tasks: 1) unique identifier creation, 2) demographic data, 3) Learning Styles Inventory, and 4) Association Rules instruction and quiz. The average amount of time for completion was 32 minutes, with a range of 19 to 44 minutes (rounded to the nearest minute). Table 4.1 provides a summary of the data collected.

Table 4.1

*Summary of Data Collected (n=86)*

Variable	Value
Average Age	21 years
Gender	
Male	64
Female	22
Academic Year	
Freshman	2
Sophomore	18
Junior	20

Senior	46
Unique Academic Majors	7
Internet Proficiency (self-rated)	
Low	0
Moderate	24
High	31
Very High	30
Highest Level College Math Completed	
100	36
200	42
300	8
400+	0
Learning Styles	
Accommodator	37
Assimilator	11
Converger	28
Diverger	10
Teaching Methodology Received	
Direct Instruction	43
Concept Attainment	43

---



Evaluation of these demographic and descriptive variables collected from the study participants reveals that they do form a representative and useful data set. The sample size is large enough to provide meaningful inferential outcomes, and the composition of the group is reflective of the desired student population.

#### *Findings for Research Question One: Models of Teaching*

Research question one asks: “Is there a significant difference in assessment scores between the Direct Instruction and Concept Attainment delivery models?” The null hypothesis for this question states that there is no statistically significant difference between the two instructional delivery methods, or in other words, one teaching methodology will not generate significantly higher quiz scores over the other. Although students who participated in the study were randomly assigned to receive one instruction method or the other by a computer generated randomization algorithm, care was taken to ensure an identical sample size for each of the two groups. The resulting groups were each comprised of 43 observations, enabling the use of a two-tailed *t*-test for the determination of statistical significance. Table 4-2 presents the descriptive statistics for the Concept Attainment participant group, while table 4-3 offers the same for the students who received their data mining instruction via Direct Instruction. Outcomes of the *t*-test follow.

Table 4.2

*Descriptive Statistics for Concept Attainment Participants (n=43)*

Statistic	Value
Mean	5.81
Median	6
Mode	5
Standard Deviation	2.11
Minimum Value	1
Maximum Value	10
Sample Variance	4.44

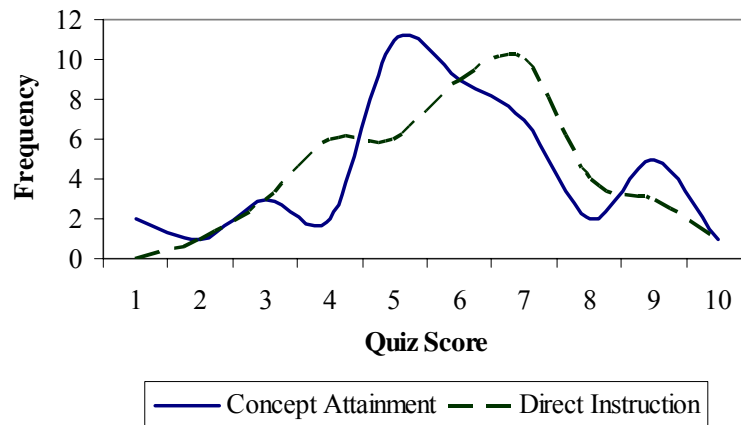
Table 4.3

*Descriptive Statistics for Direct Instruction Participants (n=43)*

Statistic	Value
Mean	6
Median	6
Mode	7
Standard Deviation	1.84
Minimum Value	2
Maximum Value	10
Sample Variance	3.38

In evaluating the measures of central tendency between the two groups, we find that they are very similar. Figure 4.4, which depicts the frequency distributions of quiz scores for each teaching methodology, further illustrates the similarities between the two groups' scores. Given the sample sizes for the two groups, the scores are fairly normally distributed.

Figure 4.1. Frequency Distributions for Quiz Scores within Models of Teaching



With these similarities in mind, the *t*-test was performed in order to determine whether or not the differences in quiz scores are statistically significant. In keeping with the norm for social science research, an alpha level of .05 ( $\alpha = .05$ ) was used. The resultant *p* value for a two-tailed test yielded .63 ( $t=-0.48$ ,  $df=42$ ,  $p<.64$ ), which is far too large to suggest statistically significant differences. Therefore, we fail to reject the null hypothesis for research question one of this dissertation.

#### *Findings for Research Question Two: Learning Styles*

Our second research question seeks the answer to the following: “Is there a significant difference in assessment scores among student Learning Styles”? Based on their inventory responses, each student was identified by one of Kolb’s four learning style categories. Tables 4.5 through 4.8 summarize the descriptive statistics for each of these four learning style groups.

Table 4.4

*Descriptive Statistics for Quiz Scores within the Accomodator  
Learning Style (n=37)*

Statistic	Value
Mean	5.84
Median	6
Mode	5
Standard Deviation	2.02
Minimum Value	1
Maximum Value	10
Sample Variance	4.08

Table 4.5

*Descriptive Statistics for Quiz Scores within the Assimilator Learning Style (n=11)*

Statistic	Value
Mean	6
Median	6
Mode	4
Standard Deviation	2
Minimum Value	3
Maximum Value	9
Sample Variance	4

Table 4.6

*Descriptive Statistics for Quiz Scores within the Converger Learning Style (n=28)*

Statistic	Value
Mean	6.11
Median	6
Mode	6
Standard Deviation	2.04
Minimum Value	1
Maximum Value	10
Sample Variance	4.17

Table 4.7

*Descriptive Statistics for Quiz Scores within the Diverger Learning Style (n=10)*

Statistic	Value
Mean	5.5
Median	6
Mode	6
Standard Deviation	2.04
Minimum Value	3
Maximum Value	8
Sample Variance	4.17

Using an ANOVA statistical test with an alpha level of .05 ( $\alpha = .05$ ), as we did in the hypothesis test for research question one, we generate a  $p$ -value of .86 (F Crit.=2.71, df=3, 82,  $p=.86$ ) for this statistical test. This  $p$ -value is again too large to indicate statistically significant differences between the four learning style groups. Thus, in response to research question two we also fail to reject the null hypothesis. No significant differences in quiz scores were identified along learning style lines.



The largest potential gap which could have resulted in significant differences between groups exists between Convergers, with a mean quiz score of 6.1, and Divergers, with a mean quiz score of 5.5. Although this gap is recognizable, further experimental analysis on these data show that with a maximum mean of 6.1, the minimum mean would have to be at or below 4.8 with an alpha of .05 in order to find significant differences in scores. Because we failed to reject our first two null hypotheses, we did not evaluate statistics for research questions three and four as part of our finds, however some discussion of these items is included in the next chapter.

#### *Experiment-wise Error Rate*

In keeping with established social science research practices, the experiment-wise error rate for Type I errors in this study is calculated as  $1-(1-.05)^2$ . This formula is derived from Moran (1986) and Sawilowsky & Kelley (1994), which set forth approaches and methodologies for controlling for error rates when multiple univariate statistical tests are done in a single study. With an alpha level of .05 ( $\alpha = .05$ ) for both the *t*-test and the ANOVA, we use this value as the confidence level in the formula, square that figure to account for the two univariate tests, and then subtract that value from one, which is representative of a 100% confidence against Type I error. This yields an experiment-wise error rate of .0975 for this study, or in other words, a 9.75% chance of having committed a Type I error. Since we failed to reject both null hypotheses tested, it is unlikely that this larger error rate has any effect on the findings of this study.

#### *A Data Mining Approach to the Data*

As an alternative to hypothesis testing, data mining offers analytical and quantitative options for locating and interpreting patterns within data mining. This type of exploratory data analysis allows us to approach a data set without any preconceived notions about what we expect

to find. Often, results of data mining are characterized by “levels of interestingness”, as opposed to more rigid statistical outcomes such as significance (Han & Kamber, 2001). Since this dissertation focuses specifically on the discipline of data mining, some time has been dedicated to mining the data set in search of interesting patterns.

The first exploration of the data takes the form of an Association Rule model, to see if any elements within the data set were found so frequently together that we might consider their proximity a rule. In creating Association Rule models, two mechanisms exist to test the frequency and reliability of the rule. These two mechanisms, the support and confidence percentages, are explained in chapter one of this dissertation, under operational definition number eight. To review, the support percentage indicates the minimum portion of records in the data set which must comprise a given rule, while the confidence percentage sets the minimum number of records containing one part the rule which must also contain the other part (or parts) of the rule. The data set was mined using the Apriori algorithm, the primary methodology for creating Association Rule models.

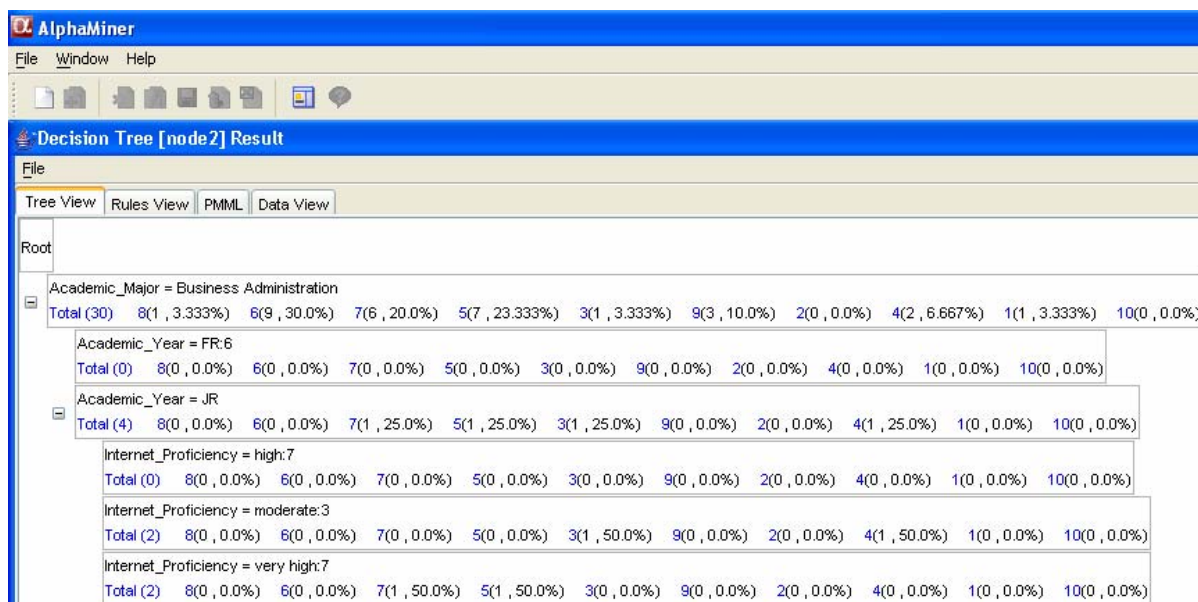
Even when setting the support and confidence percentages very low, no truly beneficial rules emerge from the data. Associations such as academic major and learning style may reveal some interesting patterns regarding courses of study and the audience they attract, however the participants did not create any such patterns in the data they generated for this study.

With no interesting results from Association Rule modeling, another data mining methodology was employed. Decision Trees are a predictive modeling approach which also generates rules, however under Decision Tree methodology the rules predict potential outcomes from past performance. It could be useful in this study to know if any of the descriptive characteristics surrounding our data are predictive of the students’ quiz scores. The variable to

be predicted by the model is referred to as the “target variable”, which is the quiz score in this study. Knowing any predictive elements of the quiz score would enable us to target those attributes which yield the highest or lowest scores, and design teaching interventions around them.

Decision Trees are so called because the predictive model is created by segmenting the data by predictive variables, from most predictive variables to least predictive variables. When generated for the data set from this study, the first predictive variable found is Academic Major. Within the academic major of *Business Administration*, for example, another predictive variable, Academic Year, is found. Further, within the Academic Year of *Junior*, we find that the variable Internet Proficiency is predictive. When viewing the tree for a *Business Administration* major who is a *Junior* and is *moderately* proficient with the Internet, our model predicts that students fitting this profile will score a three or four out of ten on the administered quiz. Figure 4.13 demonstrates this data mining result in tree form.

Figure 4.2. Decision Tree Exemplar from the Study Data



Aside from tree view, decision trees can also be displayed as a set of predictive rules. These rules can be sorted to show which rules predict the highest or lowest scores, which rules are supported by the greatest number of records, or which rules are most consistent within the data set. Table 4.8 shows the Decision Tree rules from this study's data set for the highest predictable quiz score, the lowest predictable quiz score, the prediction rule with the highest number of supporting records, and the rule with the greatest confidence percentage. We say *predictable* when referring to the quiz scores since not every possible score is necessarily predictable in the data set.

Table 4.8

*Decision Tree Rule Exemplars from the Study Data*

Rule	Supporting Records	Confidence Percent	Predicted Score
Academic_Major = Business Administration AND Academic_Year = SR AND College_Math >= 100 AND Instructional_Method = Concept Attainment	2	100%	9
Academic_Major = Business Administration AND Academic_Year = JR AND Internet_Proficiency = moderate	2	50%	3

Academic_Major = Information Technology Leadership			
AND Academic_Year = SR	12	41.7%	7
Academic_Major = Business Administration AND	6	100%	5
Academic_Year = SR AND College_Math > 100 AND			
Instructional_Method = Concept Attainment AND			
Learning_Style = Accomodator			

---

These four rules are interpreted in the following explanations:

- **Highest predicted quiz score:** If a student's major is Business Administration, their academic year is Senior, their highest level of college math is at least 100 level, and they received their data mining instruction via Concept Attainment, we predict with 100% confidence that the student will score a nine out of ten on the Association Rules quiz. This confidence is based on two records from the data set.
- **Lowest predicted quiz score:** If a student's major is Business Administration, their academic year is Junior, and their Internet proficiency is Moderate, we predict with 50% confidence that the student will score a three out of ten on the Association Rules quiz. This confidence is based on two records from the data set.
- **Greatest number of supporting records:** The most frequently supported rule in this decision tree matches 12 of the 86 observations in the data set. This rule shows that when the student's major is Information Technology Leadership and their academic year is Senior, we predict with 41.7% confidence that the student will score a seven out of ten on the Association Rules quiz.

- **Highest confidence percent:** There are six different rules in the decision tree which yielded confidence percentages of 100%. Two of the six are shown in Figure 4.14 as examples. In the bottom example, if a student's major is Business Administration, their academic year is Senior, their highest level of college math is greater than 100 level, they received their data mining instruction via Concept Attainment, and their learning style is Accomodator, we predict with 100% confidence that the student will score a five out of ten on the Association Rules quiz. This confidence is based on six records from the data set.

Having interpreted these rules, it is critically important that we understand the value of each rule in light of the supporting records and confidence percent. The highest score rule listed first in Table 4.8 is a prime example. Although the confidence percent is high (100%), the number of supporting records is low. This means that for all students who participated in this study and who fit the rule profile as interpreted above, the quiz score earned was nine—but that is only the result two people. If we were to find even a handful of other students who fit this exact profile and administer the data mining instruction and quiz to them, it is highly unlikely that they would all score a nine out of ten on the quiz. Thus, although this was true 100% of the time in this study, we would want to use caution in extending this claim of predictability very far beyond this study.

In similar fashion, we must use caution in interpreting the rule with the greatest number of supporting records. This rule is true for more than 13% of the records in the data set, however of all records where the student fits the profile, only 41.7% of the time did the student score a seven on the quiz. In other words, more than half of the time this highly supported rule will not predict the correct quiz score.

The most reliable rule found is the following:

- Academic\_Major = Business Administration AND Academic\_Year = SR  
AND College\_Math > 100 AND Instructional\_Method = Direct Instruction

This rule is supported by nine records with a confidence level of about 56% and predicts a score of seven out of ten on the quiz.

In assessing the outcome of this exercise then, we can conclude that mining the data set yields very similar results to the hypothesis testing conducted and reported in the beginning of this chapter: no definitive connections were found between learning styles, the chosen teaching methodologies, and the resultant quiz scores earned by the students.

### *Summary of Study Findings*

In testing the first two hypotheses, which sought to find significant impacts between teaching methodologies, learning styles, and data mining quiz scores, we failed to reject both null hypotheses. No significant differences were found between the Direct Instruction and Concept Attainment teaching model groups, and none were found among the four learning style groups identified through Kolb's Learning Style Inventory. This rendered hypothesis testing for research questions three and four meaningless, since these hypotheses sought to locate significant differences within more effective teaching models or learning styles.

Data mining was also conducted on the study data. No patterns which would create Rules of Association were identified. A Decision Tree generated from the data did yield results which supported the conclusions of the statistical tests. No definitive rules emerged with high enough support and confidence levels to use in decision making, however a few of the rules did provide insights into possible matches of teaching models, learning styles, and quiz scores.

## CHAPTER V: SUMMARY, DISCUSSION AND RECOMMENDATIONS

The discipline of data mining is still relatively new, having achieved a growing acceptance and practice only within the past decade. As data mining software tools and practices have become more accessible, both from an economical and intellectual standpoint, it makes sense to begin to examine the possibility of teaching the discipline outside of its traditional sphere—that of the graduate and Computer Science curriculum. This dissertation adds to a small body of scholarly literature dedicated to the pedagogy of data mining.

### Summary and Discussion

Data mining has grown out of the largest research universities and corporations in the world. This exploratory approach to data analysis has changed the nature of data collection, management and evaluation, subsequently altering the nature of decision support systems as well.

Today, data mining software is increasingly available to individuals and small organizations at low or even no cost through efforts such as the open source software movement. Many of the emerging data mining software tools run on inexpensive hardware which are also readily available to individuals and organizations of any size. Thus the current technological environment sets the stage for data mining education to expand beyond its former domain on the mainframes and multi-node networks of universities and corporations to the classrooms of undergraduate colleges and disciplines outside the historical norm.

This particular study focused on the data mining technique of Association Rules, and sought to address four research questions:



1. Is there a significant difference in Association Rules quiz scores between the teaching methodologies of Direct Instruction and Concept Attainment?
2. Is there a significant difference in Association Rules quiz among Kolb's four learning styles?
3. If one model of teaching proves to yield higher scores, are there learning styles within the more successful teaching model which generate significantly different scores from the other learning styles?
4. If one or more learning style proves to yield higher scores, does one of the models of teaching within the more successful learning style(s) generate significantly different scores from the other teaching model?

#### *Discussion of Demographic Data*

In general, findings from evaluation of the demographic data are not surprising; however there are a few items which merit some additional discussion. The disparity between male and female participants is not reflective of the overall population of the college, which is closer to a 50-50 split along gender lines. Definitive reasons for the gender distribution being different from the college's norm are unclear, and while worthy of note, investigation into potential explanations falls outside the scope of this study. Additionally, gender is not considered one of the independent variables in any of the hypotheses in this study, and as such, this disparity in gender does not compromise the validity of the outcomes reported.

Other items in the demographic data which justify some discussion include academic major, self-rated Internet proficiency, and college math levels. The major areas of study for students in this study include Arts, Humanities, and Social and Natural Sciences. This range of majors forms a broad and diverse backdrop for one of the major concerns in this dissertation, that

of teaching data mining to non-computer scientists. Certainly this participant group represents the kind of academically diverse pool envisioned at the outset.

Predictably (given the age of the students who completed this study) none of the participants rated themselves as having a low proficiency when using the Internet. This question was asked of the students largely because the instruction, whether delivered via Direct Instruction or Concept Attainment, was done over the Internet. Our objective was to ensure that this delivery medium did not have any unintended effect on the quiz scores being evaluated. A Pearson correlation was run on these two data variables (Internet proficiency and quiz score), yielding a coefficient of .10, indicating that the two have very little, if anything, to do with one another's variability. This is important to note as it lends additional support to the claim made in chapter two where we suggest, based on published literature, that the choice of online instructional delivery will not, and did not, have any undue influence on the independent variable (quiz scores) being studied.

Data mining is a math-driven, analytical, and quantitative discipline by nature. We were therefore interested in a potential correlation which might exist between the level of college math completed by the student participants and their scores on the Association Rules quiz. A Pearson Correlation performed on these two variables resulted in a coefficient of -.14, which also indicates a very low rate of shared variability. This is perceived as a positive given that our target audience is non-Computer Science students, and it is very likely that if Computer Science students had been included, the number of study participants having completed upper-level course work in mathematics would likely increase.

### *Outcomes for Models of Teaching*

In considering the failure to reject our first null hypothesis, there are several points of explanation which may shed light on the result. First, since the material taught was new for all participants, all started from a common point. Whether they embarked on new knowledge by listening and observing, or by interacting and making choices, each student needed a basic level explanation and guidance in order to understand Association Rules and become equipped to take the quiz. Additionally, the participants in the study represent a fairly homogenous group. For the most part, all students share a very similar academic, interpersonal, and socio-economic background, which in turn helps to ensure that their starting point for knowledge acquisition in data mining is also very similar. Therefore, whether beginning this learning via Direct Instruction or Concept Attainment, it is unlikely that much divergence would be found after a single lesson in Association Rules. The fact that only one data mining lesson and assessment was administered serves as a constraint for the study.

The degree of variance in each of the two instructional groups may begin to expose those participants who have an aptitude toward the subject matter and those who do not; however this is not correlated to choice of instruction methodology in this study. The level of attention and interest generated by the one instructional methodology or the other may have contributed to slightly higher scores among the more attentive group; however these higher average scores are not significant, and could also be attributed to some other factors, or simply to chance. Given that students were assigned their instruction methodology at random, it is impossible to say whether students in one group paid closer attention or were more engaged in the learning activities than students in the other group. Ultimately, we can only conclude from this portion of

the study that teaching methodology did not significantly impact these students' performance on their first data mining quiz.

It is important to note that teaching model choice should not be discarded in light of the failure to reject the first null hypothesis. Concept Attainment as a model of teaching urges students to draw upon prior knowledge in order to distinguish between exemplars and non-exemplars of the subject matter being taught. The target audience, by design, did not have much, if any, prior knowledge upon which to draw when identifying exemplars of Association Rule concepts and techniques. The Concept Attainment group of students did perform slightly lower on the quiz than did the Direct Instruction group. Since the gap was not significant however, it may well be that once equipped with a foundational knowledge of data mining approaches, Concept Attainment could prove to be a highly effective and useful method for teaching data mining knowledge to students.

In addition, we tested only two teaching methodologies. These two were chosen intentionally based upon published literature. There are other teaching approaches which may also yield interesting and perhaps even significant results which could also be tested. These models may include some of the following (Allen, 1996):

- *Mind Mapping*, a personal, generative approach to teaching which some data mining algorithms seek to emulate as artificial intelligence.
- *Memory Modeling*, a cognitive and information processing guided learning experience which may help students relate to the statistical and logical methods applied in data mining.

- *Cooperative Learning*, a social constructivist teaching model which encourages groups of learners to create synthesis and reach conclusions about the subject matter through discussion, debate and consensus.

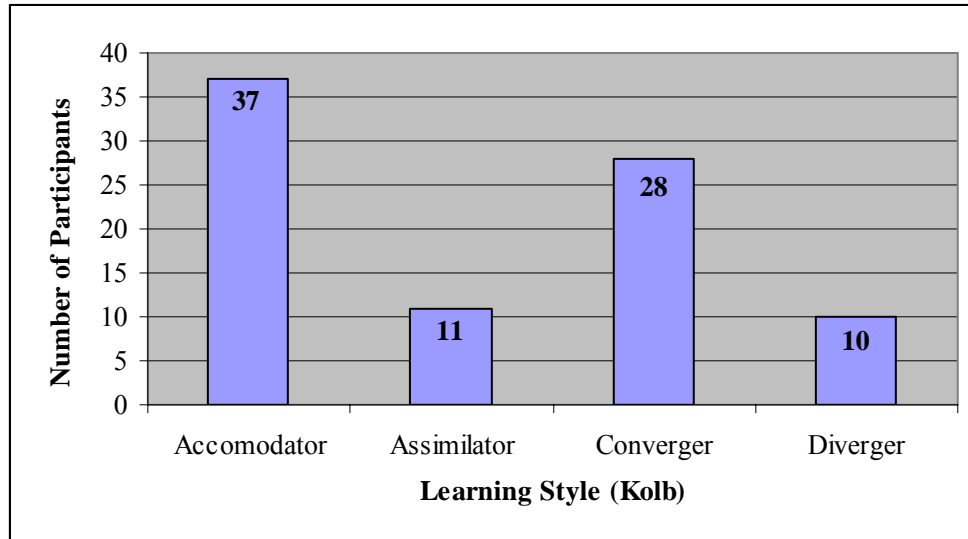
There are others which may also serve as good candidates for further testing as well. In general, it is believed that teaching choices which support good quantitative analysis skills will result in a more successful learning experience in undergraduate data mining education, and further research into these areas is justified.

### *Learning Style Outcomes*

We have also hypothesized that one or more of Kolb's Learning Styles would yield significantly difference quiz scores from the others. Based on the analysis of variation performed on the data, we failed to reject the null hypothesis for this second research question. Overall, Convergers emerged as the most successful group in terms of quiz score, with their Diverger counterparts showing as least successful on the quiz. This outcome was not statistically deemed to be different, however comparison of the means did show these two groups to be on opposite ends of the spectrum of quiz scores. This outcome is consistent with Kolb's description of his learning style groups, particularly with the Divergers, who do not like to be bound by "correct solutions" or formulae. Although data mining does not necessarily lend itself to one specific right answer every time, it is still highly quantitative and logical in nature, attributes which are sure to appeal more strongly to the Converger group.

In order to best visualize these data across all four learning style groups, figures 5.1 through 5.3 are presented below with some explanatory discussion.

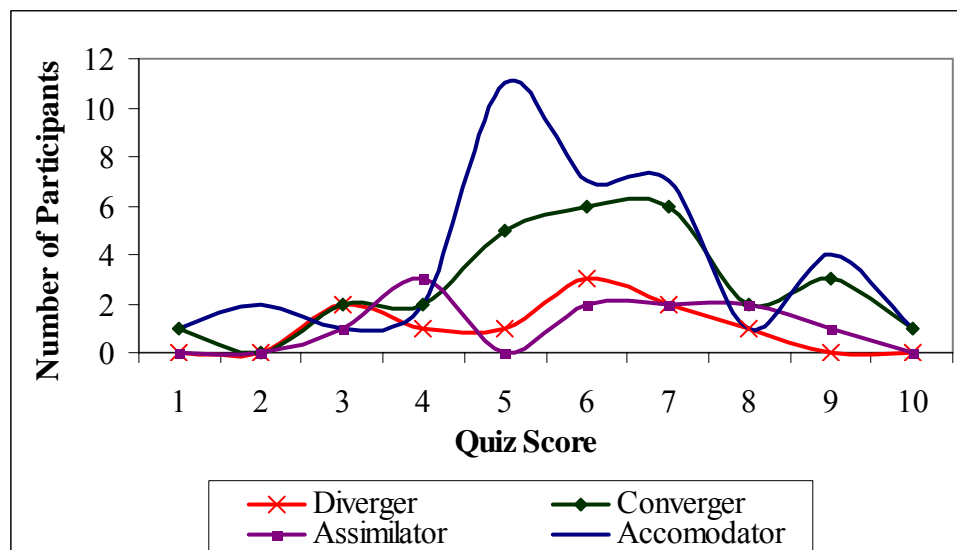
Figure 5.1. Participant group sample sizes by Learning Style



According to Choi and Washington (1988), Kolb's learning style groups do not have a consistent frequency distribution across disciplines or research studies. In their study, they found Assimilator to be the most frequent learning style among professional librarians for example, while Accomodator was the least common for this profession. Kolb (1985) notes that any given group of individuals will display a unique makeup when administered the Learning Styles Inventory, and that the external factors which would explain the group's composition are varied. Therefore, the inventory does not attempt to explain *why* a certain group is distributed the way it is across the four learning styles, it simply aids each participant in determining their own learning style. The group which chose to participate in this study was largely comprised of Accomodators and Convergers. According to Kolb's interpretation booklet, Accomodators have a tendency toward problem solving through trial and error, and prefer to involve themselves as subjects in new experiences. Convergers, on the other hand, enjoy the application of practical and methodical means for reaching conclusions, and prefer scenarios where a single, "correct"

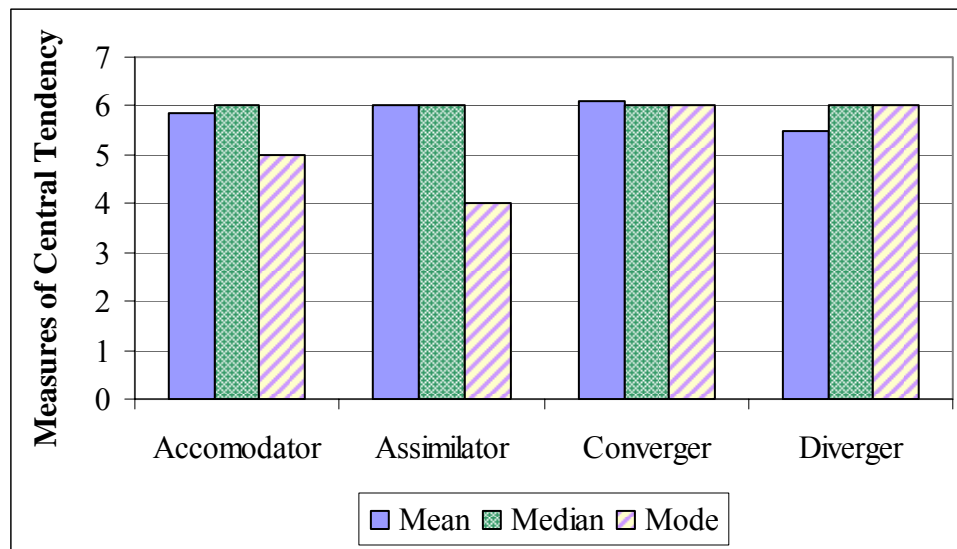
solution can be reached. Both of these learning styles would lend themselves well to the study of data mining, considering the fact that it is both an *exploratory* approach to information analytics, and a highly *methodological and quantitative* discipline. Figure 5.2 illustrates that these two learning style groups have taken on a more normal looking distribution in this study; however, this does not necessarily translate to significantly higher quiz scores than those of the other two groups.

Figure 5.2. Quiz Score Frequency Distributions by Learning Style



In interpreting Figure 5.2, it may be tempting to see the dramatic amplitude for Accomodators and Convergents illustrated at the quiz score of 5 (or a 50% on the quiz) in contrast to the corresponding drop at that quiz score by the Divergers and Assimilators. However, Figure 5.3 sheds additional light to show that these differences are more a reflection of relative sample size than of statistically higher scores.

Figure 5.3. Quiz Score Means, Medians and Modes by Learning Style



This figure more clearly illustrates the limited amount of variance from one learning style group to another while controlling for the disparity in sample sizes, suggesting that statistically significant differences between learning style groups are unlikely to be found. This turns out to be true when tested using ANOVA to evaluate the scores among the four groups.

These findings are a departure from some other published literature in computer-related education. Thomas, Ratcliffe, Woodbury and Jarman (2002) found significant differences on an Introduction to Programming test among learning style groups. This study however used the Felder-Silverman learning style assessment instrument, which groups students slightly differently than does Kolb's inventory. Still, their results seem to parallel the findings in this study. The group in their study which scored highest was described as "reflecting, sensing, verbal, sequential", which most closely matches the description of Convergers in Kolb's categories. This study, like the Thomas, et al. study, involves introductory computer-related education. Convergers did score highest on the data mining quiz among the four groups.



Similarly, the Thomas, et al. study lists their lowest scores from a learning style group described as “reflective, intuitive, visual, global”, which closely mirrors the description of Kolb’s Divergers, who also scored lowest in this study. In considering the diagram of Kolb’s Learning Styles in chapter two, Figure 2.1, the findings of both of these studies lend support to the juxtaposition of the Converger and Diverger learning styles, with the Converger appearing to possess more aptitude in computer-related subjects such as programming and data mining.

#### *Other Outcomes of Data Analysis*

The hypotheses for research questions three and four were not statistically tested, since we failed to reject the null hypotheses for research questions one and two. The data were reviewed in the context of both the interaction of learning style and model of teaching though, by simply comparing mean scores in a four by two matrix. Overall, the highest mean quiz score (6.8 out of 10) was achieved by the Assimilators who received their Association Rules lesson via Direct Instruction (see Table 5.1). Interestingly, this same group scored lowest when taught via Concept Attainment (5.2 out of 10), suggesting that at least for Assimilator-type learners, choice of teaching methodology does matter. Convergers scored highest among the Concept Attainment teaching group (6.3 out of 10), the only learning style group under this model of teaching to score above a six on average on the quiz. This may be attributable to the Convergers’ preference for quantitative and mathematical disciplines, which provides more context upon which to draw under the more constructivist approach of Concept Attainment.

Table 5.1

*Quiz Score Means by Learning Style and Teaching Model*

Learning Style	Concept Attainment Mean	Direct Instruction Mean
Accomodator	5.7	6.0
Assimilator	5.2	6.8
Converger	6.3	5.9
Diverger	5.6	5.4

Of the pairs of means shown in Figure 5.1, only the Assimilator means are statistically different from one another, verified by a  $t$ -test ( $t=-1.37$ ,  $df=4$ ,  $p<.000005$ ). This apparent relationship merits further study, and could support the idea that teaching and instructional delivery choices can help students improve their performance in subjects in which they do not necessarily have a natural aptitude. To further draw upon the example at hand, Assimilators in this study did not score the highest overall as a group, however they did score both lower and higher than any other group, solely based upon the model of teaching to which they were assigned. For this group in particular, acquisition of knowledge in the field of data mining appears to have heavily relied upon teaching choices. This finding further supports the notion that a combination of learning style and teaching model may be more effective for some learners. As an example, consider the differences between the two teaching models for the first two

learning style groups as opposed to the latter two learning style groups. Accomodators and Assimilators performed better under Direct Instruction, while Convergers and Divergers both performed better after receiving Association Rules instruction through Concept Attainment. This phenomenon is best illustrated when considering the Assimilator means, where the scores generated after Direct Instruction are clearly higher than the scores earned after Concept Attainment.

### Recommendations

The results of this study show that although there are some tendencies, there are no statistical reasons to choose Direct Instruction or Concept Attainment or target a specific learning style for introductory instruction in data mining. This adds to the small amount we know about data mining pedagogy. Perhaps more important than these findings however, is the illustration that much more scholarly research is needed in the area of introductory data mining teaching and learning.

### *Implications for Data Mining Education*

A major objective in conducting this research was accomplished by generating evidence that the elementary concepts and techniques of data mining can be effectively taught to an undergraduate non-Computer Science audience. The claim has been made that these concepts were effectively taught during this research. Some may question that claim, given the relatively low mean scores achieved by the participants, about 6 out of 10, which would equate on a letter grade scale to a low D or an F. This result should be considered in light of several facts:

- The quiz was intentionally designed to generate as much variance as possible, which is preferred for hypothesis testing. This pushed the mean score toward the center of the 0 to 10 scale.
- The participants were receiving their very first instruction in data mining, and it was delivered over a very short period time (about 30 minutes on average).
- Although students could ask questions by coming to the instructor in person, the online delivery environment did not create an interactive opportunity to ask questions of an instructor.
- Students were not compensated in any way, probably lowering their motivation to put forth their best possible effort to get every single quiz question right.

In spite of these limitations, students performed admirably on the quiz, some even scoring a perfect 10. Thus the relatively low mean should not be construed as a failure on the part of the participants to acquire and demonstrate knowledge of basic data mining content.

As educational offerings in data mining expand, a thorough understanding of effective teaching and learning within the context of the content will improve student outcomes. This dissertation serves as an early starting point in determining what truly constitutes *effective* teaching and learning in introductory data mining education to a nontraditional group of students. In order for data mining to expand beyond its historical confines and become available to a broader, more intellectually diverse audience, this type of research is desperately needed. As the software and hardware tools which enable data mining become more accessible to the masses, instruction on how to effectively use these tools will also be needed. The tools are now emerging, driven by demand for them from new audiences. These audiences know that data mining can support process management and decision making, but are not yet equipped to use

data mining effectively. By studying and finding effective ways to place these tools in the hands of the individual and small organization, data mining educators can become the change agents which will bring data mining instruction out of the research universities and large corporations, and into the hands of the new target audience. Young people earning bachelor's degrees in finance, marketing, customer service, business administration, health care, or a host of other disciplines, can leave college equipped with the knowledge and skills to apply data mining at microeconomic levels. The key at the moment rests with data mining educators, who will pursue and continue research with an eye toward shifting the current educational environment surrounding data mining.

#### *Future Study*

In this study we failed to reject two null hypotheses. These findings can serve as a catalyst for future research. Further study of this same nature could be performed using other data mining lessons as the content foundation. For example, lessons on Decision Trees, k-Means Algorithms, or Logistic Regression could be used within the same methodological framework as this study in order to further confirm or to refute the findings reported here. Although the content would be different in such studies, if similar patterns emerge from the data, conclusions could be reliably drawn about the interaction between models of teaching, learning styles, and data mining knowledge acquisition.

This study examined two specific models of teaching. Others have been recommended for study in conjunction with introductory data mining education earlier in this chapter. One could look specifically at teaching families, such as behavioral, cognitive and constructivist epistemologies as more or less effective in this teaching area. As has been discussed, mathematical and logical methods have formed the discipline of data mining; however data

mining itself need not be confined to the traditional methods of teaching used in these ancestral disciplines. Data mining has become its own, unique discipline, and learners within this area may well benefit from different instructional approaches. Only through research, perhaps even through action research, can we determine whether or not this is true and if so, where the more effective approaches are found.

Similarly, Kolb's Learning Styles are not the only widely accepted approaches to classifying learning categories or groups. Learning style inventories which identify learners as Visual, Auditory or Kinesthetic (VAK) have become widely accepted in scholarly research, and could provide a different mechanism for comparison than have Kolb's categories in this study (Tanner & Allen, 2004). It is entirely possible that if the participants in this study had been classified according to VAK learning styles, their scores may have revealed statistically significant differences. It may be interesting to perform a simplified study, wherein students complete both Kolb's Learning Styles Inventory, and a VAK learning styles assessment, and then to compare group differences on quiz or other data mining assessment scores using the two different learning style classifications.

In the findings reported in table 5.1, we have also discussed the fact that mean quiz scores were significantly different between Direct Instruction and Concept Attainment for the Assimilators in the participant group. There is no explanation for this within this study and certainly further research is justified in order to determine what, if anything, is indicated by the gap discovered within this learning style group.

Finally, it is worth noting that data mining is a highly interdisciplinary field, which affords the opportunity to fuse both quantitative and qualitative research methodologies in a research project or study. This type of approach to data analysis and interpretation can allow for

the location of interesting patterns which tell a story or support a certain position or belief. This outcome is useful for decision making which goes beyond simply determining whether or not certain relationships are statistically significant. The conversation which results from data mining can often result in discoveries from the information that were not readily apparent and may not have been identified through a standardized statistical test.

It is evident from these suggestions that there are numerous possibilities for further research in the area of introductory data mining education. Given the limited current collection of published works on data mining pedagogy, researchers interested in this area are urged to pursue work in order to further add to the body of knowledge.

### Conclusion

We live in the so-called information age. Reynolds & Reynolds, one of the largest data management information systems providers in the world and an early proponent of data mining, urges clients (and potential clients) through its slogan to “turn information into advantage”. Data mining has emerged since its earliest genesis in the late 1980s and early 1990s as a major component in the fulfillment of that goal. The software and hardware to perform data mining at micro- to macro-levels is now available, and knowledge of data mining concepts and techniques is all that is lacking in creating an environment where *anyone* can take advantage of this powerful data analysis and decision making discipline. The community of data mining educators has the potential, through research, to enable the shift from graduate, Computer Science data mining education, to undergraduate, multidisciplinary data mining offerings. The key to this shift lies in understanding the elements of teaching and learning which will enable the new, unique composition of potential data mining learners. As we understand effective teaching methodologies and their interaction with learning styles, we can tailor the lessons we teach in

elementary data mining courses in order to maximize student performance and equip learners with tangible knowledge and skills, enabling them to apply data mining effectively, regardless of their chosen vocation or profession.



## REFERENCES

- Adams, G. L. & Engelmann, S. E. (1996). *Research on direct instruction: Twenty-five years beyond Distar*. Seattle: Educational Achievement Systems.
- Agrawal, R. (2003). Hippocratic Computing. *Data Mining Review*, 13(12), 12-13.
- Allen, T. H. (1996). A Key to Instructional Models. *Humboldt State University Web Site*, [Electronic Version]. Retrieved on 8 February 2006 from <http://www.humboldt.edu/~tha1/key.html>.
- Allinson, C. W. (1990). Validity of the Learning Styles Questionnaire. *Psychological Reports*, 67, 859-66.
- Anane, R. (2001). Data Mining and Serial Documents. *Computers & the Humanities*, 35(3), 299-314.
- Baldwin, L. & Sabry, K. (2003). Learning Styles for Interactive Learning Systems. *Innovations in Education & Teaching International*, 40(4), 325-340.
- Boyle, E. A., Duffy, T., & Dunleavy, K. (2003). Learning Styles and Academic Outcome: The validity and utility of Vermunt's Inventory of Learning Styles in a British higher education setting. *British Journal of Educational Psychology*, 73(2), 267-290.
- Brachman, R. J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., & Simoudis, E. (1996). Mining Business Databases. *Communications of the ACM*, 39(11), 42-48.
- Bradley, P. S., Fayyad, U. M., & Mangasarian, O. L. (1999). Mathematical Programming for Data Mining: Formulations and Challenges. *Journal on Computing*, 11(3), 217-238.
- Bruner, J. S. (1950). Social Psychology and Group Processes. *Annual Review of Psychology*, 1, 119-151.

- Bruner, J., Goodnow, J. J., & Austin, G. A. (1967). *A Study of Thinking*. New York: Science Editions.
- Canfield, A. (1980). *Learning Styles Inventory Manual*. Ann Arbor, MI: Humanics Media.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). *CRISP-DM 1.0: Step-by-step Data Mining Guide*. Chicago:SPSS Press.
- Chen, G. D., Ou, K. L., Liu, C. C., & Liu, B. J. (2001). Intervention and Strategy Analysis for Web Group-learning. *Journal of Computer Assisted Learning*, 17(1), 58-71.
- Cheung, D. W., Hwang, H. Y., Fu, A. W., & Han, J. (2000). Efficient Rule-based Attribute-oriented Induction for Data Mining. *Journal of Intelligent Information Systems*, 15(2), 175-200.
- Chittleborough, G. D., Treagust, D. F., Mamiala, T. L., & Mocerino, M. (2005). Students' Perceptions of the Role of Models in the Process of Science and in the Process of Learning. *Research in Science & Technological Education*, 23(2), 195-212.
- Chiu, C. M. (2004). Towards a Hypermedia-enabled and Web-based Data Analysis Framework. *Journal of Information Science*, 30(1), 60-72.
- Choi, J. M. & Washington, N. (1988). Learning Styles of Academic Librarians and Implications for Professional Development. Proceedings of the Annual Conference of the American Library Association, New Orleans, LA, July 9-14, 1988.
- Chung, J. T. & Cheung, D. W. (2005). *AlphaMiner User Guide*. Hong Kong: E-Business Technology Institute.
- Clark, D. (2000). *Learning Styles, or, How we go from the unknown to the known*. [Electronic Version]. Retrieved on March 18, 2006 from:  
<http://www.nwlink.com/~donclark/hrd/styles.html>

- Claxton D. S. & Ralston, Y. (1978). *Learning Styles: The Impact on Teaching and Administration. (Report No. 10)*, Association for the Study of Higher Education: Washington D.C .
- Claxton, D. S. & Murrell, P. (1987). *Learning Styles: Implications for Improving Educational Practices. (Report No. 4)*, Association for the Study of Higher Education: Washington D.C .
- Cockerton, J., Naz, L., & Sheppard, S.. (2002). Factorial Validity and Internal Reliability of Honey and Mumford's Learning Styles Questionnaire. *Psychological Reports*, 91(2), 503-519.
- Cohen, F. (2003). Mining Data to Improve Teaching. *Educational Leadership*, 60(8), 53-56.
- Cohen, R. (1969). Conceptual Styles, Culture Conflict and Nonverbal Tests of Intelligence. *American Anthropologist*, 71, 828-856.
- Collins, A. & Stevens, A. L. (1983). A Cognitive Theory of Inquiry Teaching. In Reigeluth, C. M. (Ed.), *Instructional-Design Theories and Models: An overview of their current status*. Hillsdale, NJ: Erlbaum.
- Cornwell, J. M. & Manfredro, P. A. (1994). Kolb's Learning Style Theory Revisited. *Educational and Psychological Measurement*, 54, 317-327.
- Council of Presidents and State Board for Community College Education. (1989). The validity and usefulness of three national standardized tests for measuring the communication, computation, and critical thinking skills of Washington State college sophomores: General report. Bellingham, WA: Western Washington University Office of Publications.
- Dabbagh, N. & Bannan-Ritland, B. (2005). *Online Learning: Concepts, strategies and applications*. Upper Saddle River, NJ: Prentice Hall.

- De Ciantis, S. M. & Kirton, M. J. (1996). A Psychometric Reexamination of Kolb's Experiential Learning Cycle Construct: A separation of level, style, and process. *Educational and Psychological Measurement*, 56, 809-820.
- Dillon, J. T. (1998). Using Diverse Styles of Teaching. *Journal of Curriculum Studies*, 30(5), 503-514.
- Doolan, J., Demauro, J. S., Morse, L., Digiacomo, A., Fopeano, R., & Hampton, G. (1999). *New Jersey Comprehensive Education Framework*. Trenton, NJ: State Board of Education.
- Dowling, C., Godfrey, J. & Gyles, N. (2003). Do Hybrid Flexible Delivery Teaching Methods Improve Accounting Students' Learning Outcomes? *Accounting Education*, 12(4), 373-391.
- Dringus, L. P. & Ellis, T. (2005). Using Data Mining as a Strategy for Assessing Asynchronous Discussion Forums. *Computers & Education*, 45(1), 141-160.
- Duffy, T. M. & Jonassen, D. H. (1992) Objectivist and Constructivist Conceptions of Learning and Instruction. In *Constructivism and the Technology of Instruction: A Conversation*, Duffy & Jonassen, Eds. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Efthimiadis, E. (2000). Data Visualization in Information Retrieval and Data Mining (SIG VIS). *Proceedings of the ASIS Annual Meeting*, 37, 444-446.
- Ellis, S. S. (1979). Models of Teaching: A Solution to the Teaching Style/Learning Style Dilemma. *Educational Leadership*, 36(4), 274-277.
- Enkenberg, J. (2001). Instructional Design and Emerging Teaching Models in Higher Education. *Computers in Human Behavior*, 17(5/6), 495-506.
- Ester, M., Frommelt, A., Kriegel, H. P., & Sander, J. (2000). Spatial Data Mining: Database primitives, algorithms and efficient DBMS support. *Data Mining & Knowledge Discovery*, 4(2/3), 193-216.

- Evfimievski, A., Srikant, R., Agrawal, R., & Gehrke, J. (2004). Privacy Preserving Mining of Association Rules. *Information Systems*, 29(4), 343-364.
- Exner, F. & Bear, L. (1998). Advances in Knowledge Discovery and Data Mining. *Journal of the American Society for Information Science*, 49(4), 386-387.
- Fayyad, U. & Uthurusamy, R. (1996). Data Mining and Knowledge Discovery in Databases. *Communications of the ACM*, 39, 24-68.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Artificial Intelligence*, 17(3), 37-54.
- Fayyad, U., Smyth, P., & Piatetsky-Shapiro, G. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39, 27-34.
- Federico, P. (2000). Learning Styles and Student Attitudes Toward Various Aspects of Network-based Instruction. *Computers in Human Behavior*, 16(4), 359-79.
- Frawley, W. J., Piatetsky-Shapiro, G. & Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview. *Artificial Intelligence*, 13(3), 57-70.
- Freund, J. E. & Simon, G. A. (1997). Modern Elementary Statistics. Upper Saddle River, NJ: Prentice Hall.
- Fritz, S., Speth, C., & Barbuto, J. E. (2004). Exploring Relationships between College Students' Learning Styles and Motivation. *Psychological Reports*, 95(3), 969-974.
- Gagné, R. M. (1968). Learning hierarchies. Presidential Address of APA Division 15. *Educational Psychologist*, 6 (1), 1-9.
- Gershkoff, A. R. (2005). Multiple Methods, More Success: How to Help Students of All Learning Styles Succeed in Quantitative Political Analysis Courses. *Political Science & Politics*, 38(2), 299-304.

- Grashna, A. (1972). Observations on Relating Teaching Goals to Student Response Style and Classroom Teaching Methods. *American Psychologist*, 27, 144-147.
- Gray, P. (1997). Beyond 2000. *Information Systems Management*, 14(3), 87-91.
- Gwo-Jen, H. (2005). A Data Mining Approach to Diagnosing Student Learning Problems in Science Courses. *International Journal of Distance Education Technologies*, 3(4), 35-50.
- Han J. & Kamber, M. (2001). Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann.
- Han, J., Altman, R. B., Kumar, V., Mannila, H. & Pregibon, D. (2002). Emerging Scientific Applications in Data Mining. *Communications of the ACM*, 45(8), 54-59.
- Han, J., Hu, X., & Cercone, N. (2003). A Visualization Model of Interactive Knowledge Discovery Systems and its Implementations. *Information Visualization*, 2(2), 105-125.
- Harper, P. (2005). Combining Data Mining Tools with Health Care Models for Improved Understanding of Health Processes and Resource Utilisation. *Clinical & Investigative Medicine*, 28(6), 338-341.
- Hauer, P., Straub, C., & Wolf, S. (2005). Learning Styles of Allied Health Students Using Kolb's LSI-IIa. *Journal of Allied Health*, 34(4), 177-182.
- Hunt, D. E. (1974). Student Conceptual Level and Models of Teaching: Theoretical and Empirical Coordination of Two Models. New York: Ontario Institute for Studies in Education.
- Jakobson, G., Lafond, C., Piatetsky-Shapiro, G., & Rajinikanth, M. (1988). CALIDA: A system for integrated retrieval from multiple heterogenous databases. Proceedings of the Third International Conference on Data and Knowledge Engineering, 3-18.

- Ji, J., Liu, C., Sha, Z., & Zhong, N. (2005). Personalized Recommendation Based on a Multilevel Customer Model. *International Journal of Pattern Recognition & Artificial Intelligence*, 19(7), 895-916.
- Jonassen, D. H. (1990). Toward a Constructivist View of Instructional Technology. *Educational Technology*, 9, 32-34.
- Jonassen, D. H. (1994). Toward a Constructivist Design Model. *Educational Technology*, 4, 34-37.
- Jonassen, D. H., Bessner, K., & Yacci, M. (1993). *Structural Knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale NJ: Erlbaum.
- Jones, J. (2006). Stats: Scheffe and Tukey Tests. Richland Community College Web Site, [Electronic Version]. Retrieved on June 10, 2006 from:  
<http://www.richland.edu/james/lecture/m170/ch13-dif.html>
- Jones, S. (2002, January 14). A Crowded Talent Pool. *Crain's Chicago Business*, 25(2).
- Joyce, B. R. (1980). Learning How to Learn. *Theory Into Practice*, 19(1), 15-27.
- Joyce, B., Weil, M. & Calhoun, E. (2000). *Models of Teaching*. Boston: Pearson.
- Kayes, D. (2005). Internal Validity and Reliability of Kolb's Learning Style Inventory Version 3 (1999). *Journal of Business & Psychology*, 20(2), 249-257.
- Keefe, J. W. (1979). Learning Styles: An Overview. In *Student learning styles: Diagnosing and Prescribing Programs* (pp. 1-17). Ralston, VA: National Association of Secondary School Principals.
- Kolb, D. (1976). *Learning Styles Inventory*. Boston: McBer.
- Kolb, D. (1985). *LSI Learning Style Inventory: Self-scoring inventory and interpretation booklet*. Boston: McBer.

- Koob, J. J. & Funk, J. (2002). Learning Style Inventory: Issues of Reliability and Validity. *Research on Social Work Practice, 12*(2), 293-308.
- Kyuseok, S., Srikant, R., & Agrawal, R. (2002). High-Dimensional Similarity Joins. *IEEE Transactions on Knowledge & Data Engineering, 14*(1), 156-171.
- Lewis, J. M. (1980). Concept Attainment among Postsecondary Students: Instructional Design and Delivery. *Educational Technology, 20*(7), 5-17.
- Lieberman, L. M. (1982). Learning Principles and Teaching Models. *Journal of Learning Disabilities, 15*(8).
- Loo, R. (1996). Construct validity and classification stability of the revised Learning style inventory (LSI-1985). *Educational and Psychological Measurement, 56*, 529-536.
- Loo, R. (2002). The Distribution of Learning Styles and Types for Hard and Soft Business Majors. *Educational Psychology, 22*(3), 349-360.
- Luan, J. (2002). Data Mining and Its Applications in Higher Education. *New Directions for Institutional Research, 113*, 17-36.
- Mainemelis, C. B., Boyatzis, R. E. & Kolb, D. A. (2002). Learning Styles and Adaptive Flexibility: Testing Experiential Learning Theory. *Management Learning, 33*(1), 29-34.
- Martindale, T. & Ahern, T. (2001). The Effects of Three Web-Based Delivery Models on Undergraduate College Student Achievement. *International Journal of Educational Telecommunications, 7*(4), 379-392.
- Merrill, M. D. & Tennyson, R. (1977). *Teaching Concepts: An instructional design guide* (1<sup>st</sup> Ed.). Englewood Cliffs, NJ: Educational Technology.
- Miller, L. M. (2005). Using Learning Styles to Evaluate Computer-based Instruction. *Computers in Human Behavior, 21*(2), 287-306.



- Minoli, D. (1996). *Distance Learning Technology and Applications*. Boston: Artech House.
- Moran, J. D. (1986). Methodological Note: Let the t-Test Rest in Peace--A Note on the Control of Error Rates. *Home Economics Research Journal*, 14(4) 380-83.
- Morse, J. & Morse, S. (2002). Teaching Temperance to the 'Cookie Monster': Ethical Challenges to Data Mining and Direct Marketing. *Business & Society Review*, 107(1), 76-95.
- Mupinga, D. M., Nora, R. T. & Yaw, D. C. (2006). The Learning Styles, Expectations, and Needs of Online Students. *College Teaching*, 54(1), 185-189.
- Ng, R. T. & Han, J. (2002). CLARANS: A Method for Clustering Objects for Spatial Data Mining. *IEEE Transactions on Knowledge & Data Engineering*, 14(5), 1003-1017.
- Palmer, D. (2005). A Motivational View of Constructivist-informed Teaching. *International Journal of Science Education*, 27(15), 1853-1881.
- Piatetsky-Shapiro, G. & Jacobson, G. (1987). An Intermediate Database Language and its Rule-based Transformation to Different Database Languages. *Data & Knowledge Engineering*, 2(1), 1-30.
- Piatetsky-Shapiro, G. (1991). Knowledge Discovery in Real Databases: A report on the IJCAI-89 Workshop. *Artificial Intelligence*, 11(1), 68-70.
- Piatetsky-Shapiro, G. (1999). The Data Mining Industry Coming of Age. *IEEE Intelligent Systems & Their Applications*, 14(6), 32-34.
- Piatetsky-Shapiro, G., Matheus, C., & Smyth, P. (1994). KDD-93: Progress and challenges in knowledge discovery in databases. *Artificial Intelligence*, 15(3), 77-82.
- Potvin, J. (2004). The Business of Sharing. *CMA Management*, 78(6), 20-23.
- Rao, C. (2001). Statistics: Reflections on the Past and Visions for the Future. *Communications in Statistics: Theory & Methods*, 30(11), 2235-2258.

- Reichmann, S. (1978, September). Learning Styles: Their role in teaching evaluation and course design. *Proceedings of the 86<sup>th</sup> annual meeting of the American Psychological Association*: Toronto.
- Reigeluth, C. M. (Ed.) (1987). *Instructional theories in action: Lessons illustrating selected theories and models*. Hillsdale, NJ: Erlbaum.
- Resnick, L. B. (1981). Instructional Psychology. *Annual Review of Psychology*, 32, 659-704.
- Rosenshine, B. (1976). Recent Research on Teaching Behavior and Student Achievement. *Journal of Teacher Education*, 27, 61-64.
- Ross, J. L., Drysdale, M. T. B. & Schulz, R. A. (2001). Cognitive Learning Styles and Academic Performance in Two Postsecondary Computer Application Courses. *Journal of Research on Computing in Education*, 33(4), 400-412.
- Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data Mining Techniques for Customer Relationship Management. *Technology in Society*, 24(4), 483-502.
- Sarawagi, S., Thomas, S., & Agrawal, R. (2000). Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications. *Data Mining & Knowledge Discovery*, 4(2/3), 89-125.
- Sawilowsky, S. & Kelley, D. L. (1994). Meta-analysis and the Solomon four-group design. *Journal of Experimental Education*, 62(4), 361-376.
- Sayer, K. & Studd, R. (2006). Matching Learning Style Preferences with Suitable Delivery Methods on Textile Design Programmes. *International Journal of Technology & Design Education*, 16(2), 163-176.

- Saygin, Y. & Ulusory, O. (2002). Exploiting Data Mining Techniques for Broadcasting Data in Mobile Computing Environments. *IEEE Transactions on Knowledge and Data Engineering*, 14(6), 1387-1399.
- Scarpaci, J. L. & Fradd, S. H. (1985). Latin-Americans at the University Level: Implications for Instruction. *Journal of Multicultural Counseling and Development*, 13(4), 183-189.
- Searson, R. & Dunn, R. (2001). The Learning-Style Teaching Model. *Science and Children*, 38(5), 22-26.
- Sharp, J. E. (1997). Applying Kolb Learning Style Theory in the Communication Classroom. *Business Communication Quarterly*, 60(2), 129-134.
- Shen, R., Hart, P., Yang, F., Yang, Q., & Zhexue J. H. (2003). Data Mining and Case-Based Reasoning for Distance Learning. *International Journal of Distance Education Technologies*, 1(3), 46-58.
- Shi, Y., Peng, Y., Xu, W., & Tang, X. (2002). Data Mining via Multiple Criteria Linear Programming: Applications in Credit Card Portfolio Management. *International Journal of Information Technology & Decision Making*, 1(1), 131-151.
- Simonson, M., Smaldino, S. Albright, M. & Zvacek, S. (2003). Teaching and Learning at a Distance: Foundations of Distance Education. Upper Saddle River, NJ: Prentice Hall.
- Sims, R., Dobbs, G. & Hand, T. (2002). Enhancing Quality in Online Learning: Scaffolding Planning and Design through Proactive Evaluation. *Distance Education*, 23(2), 135-148.
- Skinner, B. F. (1986). Programmed Instruction Revisited. *Phi Delta Kappan*, 68(2), 103-110.
- Soukup, T. & Davidson, I. (2002). Visual Data Mining. New York: Wiley.
- Stallman, R. (2004). The Rise of the Free Software Movement. *Multinational Monitor*, pp. 29-33.

- Stein, M. L., Carnine, D. & Dixon, R. C. (1998). Direct Instruction: Integrating curriculum design and effective teaching practice. *Intervention in School and Clinic*, 33(4), 227-234.
- Stensrud, R. & Stensrud, K. (1983). Teaching Styles and Learning Styles of Public School Teachers. *Perceptual and Motor Skills*, 56, 414.
- Swanson, L. J. (1995). Learning Styles: A Review of Literature. U.S. Department of Education: Washington D.C .
- Tan, P. N., Steinbach, M. & Kumar, V. (2005). Introduction to Data Mining. Boston: Pearson.
- Tanner, K. & Allen, D. (2004). Approaches to Biology Teaching and Learning: Learning Styles and the Problem of Instructional Selection—Engaging All Students in Science Courses. *Cell Biology Education*, 3(4), 197–201.
- Thomas, L., Ratcliffe, M., Woodbury, J. & Jarman, E. (2002). Learning Styles and Performance in the Introductory Programming Sequence. *Bulletin of the Association for Computing Machinery—Special Interest Group for Computer Science Education*, 34(1), 33 - 37.
- Thompson, T. C. (1997). Learning Styles and Teaching Styles: Who Should Adapt to Whom? *Business Communication Quarterly*, 60(2), 125-127.
- Tsantis, L. & Castellani, J. (2001). Enhancing Learning Environments through Solution-based Knowledge Discovery Tools: Forecasting for self-perpetuating systemic reform. *Journal of Special Education Technology*, 16(4), 39-52.
- Tung, A. K., Hongjun, L., Han, J., & Feng, L. (2003). Efficient Mining of Intertransaction Association Rules. *IEEE Transactions on Knowledge & Data Engineering*, 15(1), 43-56.
- Urbancic, T., Skrjanc, M. & Flach, P. (2002). Web-based Analysis of Data Mining and Decision Support Education. *AI Communications*, 15(4), 199-204.
- Waldrop, M. M. (2001). Data Mining. *Technology Review*, 104(1), 101-102.

- Wautier, G. & Westman, A. S. (1995). Relationships between Learning Styles and Solutions Based on Analogies or Background Knowledge. *Psychological Reports*, 77, 1115-1120.
- Westman, A. S. (1993). Learning Styles are Content Specific and Probably Influenced by Content Areas Studied. *Psychological Reports*, 73, 512-14.
- Wilson B. G. & Cole, P. (1996). Cognitive Teaching Models. In Jonassen, D. H. (Ed.), *Handbook of Instructional Technology*, New York: Scholastic.
- Ying, A. T., Murphy, G. C., & Ng, R. (2004). Predicting Source Code Changes by Mining Change History. *IEEE Transactions on Software Engineering*, 30(9), 574-586.
- Zimmermann, T., Weibgerber, P., Diehi, S. & Zeller, A. (2005). Mining Version Histories to Guide Software Changes. *IEEE Transactions on Software Engineering*, 31(6), 429-445.
- Zaiane, O. R. (2006). A Brief History of Data Mining Research. University of Alberta Web Site, [Electronic Version]. Retrieved on February 10, 2006 from:  
<http://www.cs.ualberta.ca/%7Ezaiane/courses/cmput690/slides/Chapter1/sld011.htm>
- Zirkle, C. (2001). Access Barriers in Distance Education. *Contemporary Education*, 72(2), 39-42.

## APPENDICIES

### *Appendix A: IRB Approval Forms*

Institutional Review Board approvals from West Virginia University and Washington & Jefferson College provide support for the use of human subjects in this study.



West Virginia University

College of Human Resources and Education

June 16, 2006

MEMORANDUM

TO: Matthew North

FROM: Lynn Cartwright  
Interim Associate Dean

RE: Human Resources & Education H.S. #2006-036

**Title: "The Effect of Student Self-described Learning Styles  
Within Divergent Teaching Models in an Introductory  
Data Mining Course"**

Your Application for Exemption for the above-captioned research project has been reviewed under the Human Subjects Policies and has been approved.

This exemption will remain in effect on the condition that the research is carried out exactly as described in the application.

Best wishes for the success of your research.

cc: Deans Office  
Student Advising and Records  
Terence Ahern, Advisor

Office of the Dean

Phone: 304-293-5703  
Fax: 304-293-7565

802 Allen Hall  
PO Box 6122  
Morgantown, WV 26506-6122

Equal Opportunity/Affirmative Act



To: Matthew North

From: Dr. Lynn A. Wilson

Re: Washington & Jefferson College IRB Approval

Dear Matt:

Your research project, *The Effect of Student Self-described Learning Styles within Two Teaching Models in an Introductory Data Mining Course*, is approved as an exempt research study using human subjects at Washington & Jefferson College. It is expected that you will carry out the study exactly as described in your application. We wish you good luck in your study.

Regards,

Lynn A. Wilson, Ph.D.

Chair, W&J IRB



*Appendix B: Kolb's Learning Styles Inventory (Kolb, 1985)*

**Question 1:** I prefer

- a) hands-on learning experiences.
- b) learning through thinking and reasoning.

**Question 3:** I tend to

- a) jump right in and do something new.
- b) think about possible outcomes before trying something new.

**Question 5:** I learn best through

- a) active involvement in projects.
- b) observation.

**Question 7:** I tend to

- a) rely on feelings when making decisions.
- b) rely on logical reasoning when making decisions.

**Question 9:** I am best at learning

- a) facts.
- b) concepts.

**Question 11:** I learn best

- a) by doing.
- b) watching and then reflecting.

**Question 2:** I learn well by

- a) practical experience.
- b) applying theories to hypothetical situations.

**Question 4:** I learn more effectively from

- a) my peers.
- b) my teachers.

**Question 6:** I like learning through

- a) simulations.
- b) lectures.

**Question 8:** I would rather

- a) do volunteer work with disadvantaged youth.
- b) read about disadvantaged youth.

**Question 10:** I learn well through

- a) participating in a discussion.
- b) listening to what others have to say.

**Question 12:** I prefer assignments that

- a) require me to work examples.
- b) require me to think about situations.

*Appendix C: Direct Instruction Assessment Instrument*

**Question 1:** In creating a Rule Association model, which activity is done immediately before setting the association parameters?

- a) outlier handling.
- b) variable reduction.
- c) transactionalization.
- d) binerization.

**Question 3:** Which of the following components in a Rule Association model dictates the minimum number of observations required in order for a group of frequently matched items to be considered a rule?

- a) the support percent.
- b) the confidence percent.
- c) the minor threshold.
- d) the rule minimum.

**Question 5:** Which of the following values indicates the minimum number of observations in the data set which, if it contains one element of a rule, must also contain the other part(s) of

**Question 2:** Variables used in a Rule Association model must be of what type?

- a) continuous.
- b) categorical.
- c) alphabetized.
- d) numeric.

**Question 4:** The purpose of transactionalization in Rule Association data mining is to...

- a) determine the total number of transactions in the data set.
- b) identify the most common items in each transaction.
- c) identify all of the unique items in the data set.
- d) identify which items were purchased together in each transaction.

**Question 6:** In preparing data for Rule Association mining, all variables which are not part of each individual transaction should be set to...

the rule in order for the rule to be considered valid?

- a) the support percent.
- b) the confidence percent.
- c) the minor threshold.
- d) the rule minimum.

**Question 7:** In a marketing scenario, the purpose of Rule Association data mining would be to find...

- a) the probability that each item in the data set will be sold.
- b) the frequency with which each item in the data has been sold in the past.
- c) the frequency with which any number of items were sold together in the past.
- d) the probability that any number of items will be sold together.

**Question 9:** The industry standard algorithm upon which data mining Association Rule models are most frequently built is...

- a) Apriori.
- b) Neural Net.

- a) missing.
- b) the mean.
- c) the mode.
- d) none of the above.

**Question 8:** Imagine 1,000 grocery store receipts. We find that on at least 500 of them, whenever beer is sold, peanuts are also sold. We also find that in all, beer appears on 718 of the receipts. In order for the Association Rule (BEER => PEANUTS) to be returned as a valid rule, at what levels must our support and confidence percentages be set?

- a) 50% and 12%.
- b) 25% and 18%.
- c) 25% and 82%.
- d) 50% and 69%.

**Question 10:** In mining a data set of one million observations, an Association Rule with a confidence percentage of 89% but a support percent of only 2% is...

- a) invalid because the frequency of the rule is

c) Bayesian.

too low.

d) Squared Frequencies.

b) valid because although the association appears relatively infrequently, it is very strong.

c) invalid because the disparity between strength and frequency of the association is too great.

d) valid because although it is weak, the frequency of the association is high.

[illegible]

```

        <td>&nbsp;&nbsp;&nbsp;&nbsp;</td>
    </tr>
    <form name="AddPID" action="sqlAddPID.php" method="post">
    <tr>
        <td></td>
        <td>Please enter your Personal ID and then click "Begin the Study!" &nbsp;&nbsp;&nbsp;<input
name="PID" type="text"> <input type="submit" value="Begin the Study"></td>
    </tr>
    </form>
</table>
<strong>Inserting your Personal ID, please wait...</strong>
<?php
    $the_date = date('m-j-Y, H:i:s');
    $db = mysql_connect("localhost", "root", "");
    mysql_select_db("DissData");
    $result = mysql_query("insert into tblStudyData (Record_Date, PID) values ('$the_date',
'$PID')");

    $RID_query = mysql_query("select max(Record_ID) from tblStudyData where PID =
'$PID')");
    while($row = mysql_fetch_array($RID_query, MYSQL_NUM))
    {
        echo "<p><strong>Success! Proceeding to Step 2: Demographic
Information</strong></p>";
        echo "        echo($row[0]);
        echo ">";
    }
    ?>
<table width="100%" cellpadding="10" cellspacing="10">
    <tr>
        <td colspan="3" align="center"><h2>Demographic Data</h2></td>
    </tr>
    <tr>
        <td valign="top">
            <table border="1" bgcolor="66FF99" bordercolor="000000" bordercolordark="000000"
bordercolorlight="000000">
                <tr>
                    <td nowrap><p>Progress...</p>
                    <ol>
                        <li>Register Personal ID (Done)</li>
                        <li>Demographic Data (In Process)</li>
                        <li>Learning Styles (Pending)</li>
                        <li>Rules of Association (Pending)</li>
                    </ol>
                </td>
            </tr>
        </tr>
    </table>

```

```

        </table>
    </td>
    <form name="demographics" action="sqlAddDemographics.php" method="post">
    <?php
        echo "<input type='hidden' name='RID' value='";
        echo "$RID";
        echo ">";
    ?>
    <td valign="top">
        <table>
        <tr>
            <td>Age</td>
            <td><select name="age">
                <option value="17">17</option>
                <option value="18">18</option>
                <option value="19">19</option>
                <option value="20">20</option>
                <option value="21">21</option>
                <option value="22">22</option>
                <option value="23">23</option>
                <option value="24">24</option>
                <option value="25">25</option>
                <option value="26">26+</option>
            </select>
            </td>
        </tr>
        <tr>
            <td>Gender</td>
            <td><select name="sex">
                <option value="F">Female</option>
                <option value="M">Male</option>
            </select>
            </td>
        </tr>
        <tr>
            <td>Academic Year</td>
            <td><select name="grade">
                <option value="FR">Freshman</option>
                <option value="SO">Sophomore</option>
                <option value="JR">Junior</option>
                <option value="SR">Senior</option>
            </select>
            </td>
        </tr>
        <tr>
            <td>Academic Major</td>

```







```

        <li>Rules of Association (Pending)</li>
    </ol>
</td>
</tr>
</table>
</td>
<td valign="top">
<form name="learningstyles" action="sqlAddLearningStyles.php" method="post">
<?php
    echo "<input type='hidden' name='RID' value='";
    echo "$RID";
    echo ">";
?>
    <table border="0">
        <tr>
            <td>Please answer the following questions as thoughtfully as you can and then click
"Save My Learning Style".<hr width="350" align="left"></td>
        </tr>
        <tr>
            <td class="questions"><p>
                &nbsp; I prefer <br>
                <input name="LSI1" type="radio" value="CE"> hands-on learning
experiences.<br>
                <input name="LSI1" type="radio" value="AC"> learning through thinking and
reasoning.<hr width="350" align="left"></p>
            </td>
        </tr>
        <tr>
            <td class="questions"><p>
                &nbsp; I learn well by<br>
                <input name="LSI5" type="radio" value="CE"> practical experience.<br>
                <input name="LSI5" type="radio" value="AC"> applying theories to hypothetical
situations.<hr width="350" align="left"></p>
            </td>
        </tr>
        <tr>
            <td class="questions"><p>
                &nbsp; I tend to<br>
                <input name="LSI11" type="radio" value="AE"> jump right in and do something
new.<br>
                <input name="LSI11" type="radio" value="RO"> think about possible outcomes
before trying something new.<hr width="350" align="left"></p>
            </td>
        </tr>
        <tr>
            <td class="questions"><p>

```

<p> <input type="radio"/> I learn more effectively from  <input type="radio" value="CE"/> my peers.  <input type="radio" value="AC"/> my teachers. </p>
<p> <input type="radio"/> I learn best through  <input type="radio" value="AE"/> active involvement in  <input type="radio" value="RO"/> observation. </p>
<p> <input type="radio"/> I like learning through  <input type="radio" value="CE"/> simulations.  <input type="radio" value="AC"/> lectures. </p>
<p> <input type="radio"/> I tend to  <input type="radio" value="CE"/> rely on feelings when making  <input type="radio" value="AC"/> rely on logical reasoning when making decisions. </p>
<p> <input type="radio"/> I would rather  <input type="radio" value="AE"/> do volunteer work with  <input type="radio" value="RO"/> read about disadvantaged youth. </p>
<p> <input type="radio"/> I am best at learning  <input type="radio" value="CE"/> facts.  <input type="radio" value="AC"/> concepts. </p>

```

align="left"></p>
    </td>
</tr>
<tr>
    <td class="questions"><p>
        &nbsp; I learn well through<br>
        <input name="LSI10" type="radio" value="AE"> participating in a discussion.<br>
        <input name="LSI10" type="radio" value="RO"> listening to what others have to
say.<hr width="350" align="left"></p>
    </td>
</tr>
<tr>
    <td class="questions"><p>
        &nbsp; I learn best <br>
        <input name="LSI12" type="radio" value="AE"> by doing.<br>
        <input name="LSI12" type="radio" value="RO"> watching and then reflecting.<hr
width="350" align="left"></p>
    </td>
</tr>
<tr>
    <td class="questions"><p>
        &nbsp; I prefer assignments that <br>
        <input name="LSI9" type="radio" value="AE"> require me to work examples.<br>
        <input name="LSI9" type="radio" value="RO"> require me to think about
situations.<hr width="350" align="left"></p>
    </td>
</tr>
</table>
</td>
<td>&nbsp;&nbsp;&nbsp;&nbsp;</td>
</tr>
<tr>
    <td></td>
    <td align="center"><input type="submit" value="Save My Learning Style"></td>
</tr>
</table>
</form>
<strong>Inserting your learning styles data, please wait...</strong>
<?php
    $db = mysql_connect("localhost", "root", "");
    mysql_select_db("DissData");
    $result = mysql_query("Update tblStudyData set LSI_1 = '$LSI1', LSI_2 = '$LSI2',
LSI_3 = '$LSI3', LSI_4 = '$LSI4', LSI_5 = '$LSI5', LSI_6 = '$LSI6', LSI_7 = '$LSI7', LSI_8 =
'$LSI8', LSI_9 = '$LSI9', LSI_10 = '$LSI10', LSI_11 = '$LSI11', LSI_12 = '$LSI12' where
Record_ID = '$RID'");
    echo "<p><strong>Success! Proceeding to Step 4: Rules of Association</strong></p>

```

```

        echo($RID);
    echo ">";
?>
<table width="100%" cellpadding="10" cellspacing="10">
    <tr>
        <td colspan="3" align="center"><h2>Data Mining Lesson 1: Association Rules</h2></td>
    </tr>
    <tr>
        <td class="prograss" valign="top">
            <table border="1" bgcolor="66FF99" bordercolor="000000" bordercolordark="000000"
bordercolorlight="000000">
                <tr>
                    <td nowrap><p>Progress...</p>
                        <ol>
                            <li>Register Personal ID (Done)</li>
                            <li>Demographic Data (Done)</li>
                            <li>Learning Styles (Done)</li>
                            <li>Rules of Association (In Process)</li>
                        </ol>
                    </td>
                </tr>
            </table>
        </td>
        <td valign="top" colspan="2">
            <table>
                <tr>
                    <td><p>Association Rules are the process by which we find patterns of similar
behavior in large sets of data. For example, everyday, hundreds of grocery store receipts are
generated as cashiers scan and bag groceries. What items do people buy most often at the same
time? If we knew, could we begin to better understand shopping behaviors and use that new
understanding to improve marketing or inventory management? </p>
                        <p>Click the "Start the Lesson" link below to begin a multimedia demonstration
which will teach you how to build and interpret Association Rule models in data mining. Feel
free to take notes. When the presentation is finished, click the "Take the Quiz" button to
complete a 10-question quiz and then continue to the second lesson in this study.</p></td>
                    </tr>
                    <tr>
                        <td align="center">&nbsp;<p><a href="AssociationRules.ppt">Start the
Lesson</a></p></td>
                    </tr>
                </table>
            </td>
        </tr>
    </table>
    <?php
        <a href=http://mail.yahoo.com/config/login?/"DecisionTrees.ppt">Start the
Activity</a></p>
    </td>

```

```

        </tr>
    </tr>
    echo "<form name='conceptattainment'
action=http://mail.yahoo.com/config/login?/'ConceptAttainmentQuiz.php?RID=";
    echo "$RID";
    echo "' method='post'>";
    echo "<input type='hidden' name='RID' value=";
    echo "$RID";
    echo ">";
?>

    OR
    echo "<form name='directinstruction' action='DirectInstructionQuiz.php?RID=";
    echo "$RID";
    echo "' method='post'>";
    echo "<input type='hidden' name='RID' value=";
    echo "$RID";
    echo ">";
?>
<tr>
    <td align="center">&nbsp;<p><input type="submit" value="Take the Quiz!"></p></td>
</tr>
</form>
</table>
<strong>Inserting your quiz answers for Rules of Association, please wait...</strong>
<?php
    $db = mysql_connect("localhost", "root", "");
    mysql_select_db("DissData");
    $result = mysql_query("Update tblStudyData set DI_1 = '$DI1', DI_2 = '$DI2', DI_3 =
'$DI3', DI_4 = '$DI4', DI_5 = '$DI5', DI_6 = '$DI6', DI_7 = '$DI7', DI_8 = '$DI8', DI_9 = '$DI9',
DI_10 = '$DI10' where Record_ID = '$RID'");
    echo($RID);
    echo ">";
?> <tr>
    <td align="center">&nbsp;<p><input type="submit" value="Take the Quiz!"></p></td>
</tr>
</form>
</table>
<table width="100%" cellpadding="10" cellspacing="10">
    <tr>
        <td colspan="3" align="center"><h2>Data Mining Lesson 1: Association Rules
Quiz</h2></td>
    </tr>
    <tr>
        <td class="progress" valign="top">
        <form name="directinstruction" action="sqlAddDIQuiz.php" method="post">

```

```

<?php
    echo "<input type='hidden' name='RID' value=''";
    echo "$RID";
    echo ">";
?>
    <table border="1" bgcolor="66FF99" bordercolor="000000" bordercolordark="000000"
bordercolorlight="000000">
        <tr>
            <td nowrap><p>Progress...</p>
                <ol>
                    <li>Register Personal ID (Done)</li>
                    <li>Demographic Data (Done)</li>
                    <li>Learning Styles (Done)</li>
                    <li>Rules of Association (In Process)</li>
                </ol>
            </td>
        </tr>
    </table>
</td>
    <td valign="top" colspan="2">
        <table>
            <tr>
                <table border="0">
                    <tr>
                        <td>Please answer the following questions pertaining to Rules of Association and
then click "Save Quiz Answers".<hr width="350" align="left"></td>
                    </tr>
                    <tr>
                        <td class="questions"><p>
                            In creating a Rule Association model, which activity is done immediately before
setting the association parameters?<br>
                            <input name="DI1" type="radio" value="A"> outlier handling.<br>
                            <input name="DI1" type="radio" value="B"> variable reduction.<br>
                            <input name="DI1" type="radio" value="C"> transactionalization.<br>
                            <input name="DI1" type="radio" value="D"> binerization.<hr width="350"
align="left"></p>
                        </td>
                    </tr>
                    <tr>
                        <td class="questions"><p>
                            Variables used in a Rule Association model must be of what type?<br>
                            <input name="DI2" type="radio" value="A"> continuous.<br>
                            <input name="DI2" type="radio" value="B"> categorical.<br>
                            <input name="DI2" type="radio" value="C"> alphabetized.<br>
                            <input name="DI2" type="radio" value="D"> numeric.<hr width="350"
align="left"></p>
                    </tr>
                </table>
            </tr>
        </table>
    </td>

```

<p>Which of the following components in a Rule Association model dictates the minimum number of observations required in order for a group of frequently matched items to be considered a rule?</p> <p> <input type="radio"/> the support percent.  <input type="radio"/> the confidence percent.  <input type="radio"/> the minor threshold.  <input type="radio"/> the rule minimum. </p>
<p>The purpose of transactionalization in Rule Association data mining is to...</p> <p> <input type="radio"/> determine the total number of transactions in the data set.  <input type="radio"/> identify the most common items in each transaction.  <input type="radio"/> identify all of the unique items in the data set.  <input type="radio"/> identify which items were purchased together in each transaction. </p>
<p>Which of the following values indicates the minimum number of observations in the data set which, if it contains one element of a rule, must also contain the other part(s) of the rule in order for the rule to be considered valid?</p> <p> <input type="radio"/> the support percent.  <input type="radio"/> the confidence percent.  <input type="radio"/> the minor threshold.  <input type="radio"/> the rule minimum. </p>
<p>In preparing data for Rule Association mining, all variables which are not part of each individual transaction should be set to...</p> <p> <input type="radio"/> missing.  <input type="radio"/> the mean.  <input type="radio"/> the mode. </p>



☐ none of the above.

---

In a marketing scenario, the purpose of Rule Association data mining would be to find...

☐ the probability that each item in the data set will be sold.

☐ the frequency with which each item in the data has been sold in the past.

☐ the frequency with which any number of items were sold together in the past.

☐ the probability that any number of items will be sold together.

---

Imagine 1,000 grocery store receipts. We find that on at least 500 of them, whenever beer is sold, peanuts are also sold. We also find that in all, beer appears on 718 of the receipts. In order for the Association Rule (BEER => PEANUTS) to be returned as a valid rule, at what levels must our support and confidence percentages be set?

☐ 50% and 12%.

☐ 25% and 18%.

☐ 25% and 82%.

☐ 50% and 69%.

---

The industry standard algorithm upon which data mining Association Rule models are most frequently built is...

☐ Apriori.

☐ Neural Net.

☐ Bayesian.

☐ Squared Frequencies.

---

In mining a data set of one million observations, an Association Rule with a confidence percentage of 89% but a support percent of only 2% is...

```

        <input name="DI10" type="radio" value="A"> invalid because the frequency of the
rule is too low.<br>
        <input name="DI10" type="radio" value="B"> valid because although the
association appears relatively infrequently, it is very strong.<br>
        <input name="DI10" type="radio" value="C"> invalid because the disparity
between strength and frequency of the association is too great.<br>
        <input name="DI10" type="radio" value="D"> valid because although it is weak,
the frequency of the association is high.<hr width="350" align="left"></p>
    </td>
</tr>
</table>
<tr>
    <td></td>
    <td align="center"><input type="submit" value="Save Quiz Answers"></td>
</tr>
</form>
</table>
<XBODY>
</html>

```

# Appendix E: Complete Participant Data Set

Data collected from all student participants is included below.

Table D.1.

## Complete Participant Data Set

<u>Record ID</u>	<u>Age</u>	<u>Gender</u>	<u>Academic Year</u>	<u>Academic Major</u>	<u>Internet Proficiency</u>	<u>College Math</u>	<u>Learning Style</u>	<u>Instructional Method</u>	<u>Quiz Score</u>
1	20	M	SO	Information Technology Leadership	very high	100	Assimilator	Direct Instruction	8
2	20	F	JR	Information Technology Leadership	high	200	Converger	Concept Attainment	6
3	21	M	SR	Business Administration	moderate	200	Assimilator	Direct Instruction	7
4	22	F	SR	Business Administration	high	200	Accomodator	Concept Attainment	5
5	21	M	SR	Business Administration	moderate	200	Converger	Concept Attainment	6
6	21	M	JR	Business Administration	moderate	200	Assimilator	Concept Attainment	3
7	21	F	SR	English Business	moderate	100	Accomodator	Direct Instruction	6
8	22	F	SR	Business Administration	high	200	Diverger	Direct Instruction	5
9	21	M	SO	Information Technology	high	100	Diverger	Concept Attainment	3

10	21	M	FR	Leadership Information Technology Leadership Information Technology	high	100	Accommodator	Concept Attainment	9
11	20	F	JR	Leadership Information Technology Leadership Information Technology	high	200	Converger	Direct Instruction	5
12	20	M	SO	Leadership Information Technology Leadership Information Technology	very high	100	Assimilator	Direct Instruction Concept	9
13	22	M	SR	Psychology Information Technology Leadership Information Technology	moderate	200	Accommodator	Concept Attainment	2
14	23	M	SR	Leadership Information Technology Leadership Information Technology	very high	100	Accommodator	Direct Instruction Concept	5
15	21	F	SR	English Information Technology Leadership Information Technology	moderate	100	Assimilator	Attainment	4
16	20	M	SO	Leadership Information Technology Leadership Information Technology	high	100	Accommodator	Direct Instruction	7
17	22	F	SR	Business Administration Information Technology Leadership Information Technology	high	200	Accommodator	Direct Instruction	6
18	21	F	JR	Business Administration Information Technology Leadership Information Technology	very high	200	Converger	Direct Instruction Concept	4
19	22	F	SR	Business Administration Information Technology Leadership Information Technology	high	200	Diverger	Concept Attainment	6
20	22	M	SR	Business Administration Information Technology Leadership Information Technology	high	300	Assimilator	Concept Attainment Direct	6
21	22	F	SR	Art Business Administration Information Technology Leadership Information Technology	very high	200	Accommodator	Instruction	2
22	22	M	SR	Business Administration Information Technology Leadership Information Technology	moderate	100	Converger	Concept Attainment	9

23	19	F	SO	Information Technology Leadership	very high	100	Converger	Concept Attainment	7
24	21	F	JR	Information Technology Leadership	very high	200	Converger	Concept Attainment Direct	5
25	22	M	SR	Psychology Information Technology	moderate	200	Accomodator	Instruction	6
26	20	M	SO	Leadership	very high	200	Accomodator	Direct Instruction	9
27	24	M	JR	Thematic Major	moderate	100	Diverger	Direct Instruction	4
28	22	M	SR	Accounting Business	very high	300	Accomodator	Direct Instruction	4
29	22	M	SR	Administration Information Technology	high	100	Accomodator	Concept Attainment	9
30	21	F	SR	Leadership Business	high	200	Converger	Direct Instruction Concept	4
31	21	M	SR	Administration	moderate	200	Diverger	Concept Attainment	6
32	22	M	SR	Accounting Information Technology	very high	300	Accomodator	Concept Attainment	1
33	22	M	SR	Leadership Business	very high	100	Accomodator	Direct Instruction	7
34	21	M	JR	Administration Business	very high	200	Converger	Direct Instruction Concept	5
35	22	M	SR	Administration Information Technology	moderate	300	Diverger	Concept Attainment Direct	6
36	21	M	JR	Technology	high	200	Diverger	Instruction	8

37	21	M	SR	Leadership Business Administration	moderate	200	Converger	Direct Instruction	7
38	22	M	SR	Business Administration	high	100	Accomodator	Direct Instruction Concept	6
39	21	M	JR	English Information Technology	very high	200	Converger	Attainment	5
40	22	M	SR	Leadership Business	very high	200	Converger	Direct Instruction Direct	3
41	21	M	JR	Administration Information Technology	moderate	200	Assimilator	Instruction	4
42	20	M	SO	Leadership Information Technology	very high	200	Accomodator	Concept Attainment	9
43	22	M	SR	Leadership Business	high	300	Converger	Concept Attainment Concept	8
44	22	F	SR	Administration Information Technology	high	200	Accomodator	Attainment	5
45	23	M	SR	Leadership Information Technology	very high	100	Diverger	Direct Instruction	7
46	21	M	JR	Leadership Business	very high	100	Accomodator	Concept Attainment Concept	6
47	22	M	SR	Administration Information Technology	moderate	200	Converger	Attainment	5
48	21	M	FR	Leadership Business	high	100	Accomodator	Direct Instruction Direct	10
49	22	F	SR	Administration	moderate	200	Accomodator	Instruction	7

50	22	M	SR	Business Administration	moderate	300	Accomoderator	Direct Instruction	5
51	22	M	SR	Business Administration	moderate	200	Converger	Direct Instruction	7
52	21	M	JR	Information Technology Leadership	very high	100	Diverger	Direct Instruction	3
53	20	M	SO	Business Administration	very high	200	Converger	Concept Attainment	9
54	22	M	SR	Information Technology Leadership	very high	200	Converger	Concept Attainment	7
55	23	M	SR	Information Technology Leadership	very high	100	Accomoderator	Concept Attainment	5
56	22	F	SR	Art	very high	200	Accomoderator	Concept Attainment	5
57	21	M	SO	Information Technology Leadership	high	100	Accomoderator	Direct Instruction	5
58	23	M	SR	Information Technology Leadership	very high	100	Accomoderator	Concept Attainment	7
59	22	F	SR	Business Administration	moderate	200	Accomoderator	Concept Attainment	5
60	20	M	SO	Information Technology Leadership	high	100	Accomoderator	Direct Instruction	8
61	22	M	SR	Business Administration	high	300	Accomoderator	Direct Instruction	7
62	20	M	SO	Information Technology Leadership	high	100	Converger	Direct Instruction	6

63	20	M	SO	Information Technology Leadership	high	100	Accommodator	Concept Attainment	5
64	22	M	SR	Information Technology Leadership	very high	100	Accommodator	Concept Attainment	6
65	20	M	SO	Information Technology Leadership	high	100	Converger	Concept Attainment	3
66	20	M	SO	Information Technology Leadership	very high	100	Assimilator	Concept Attainment	7
67	20	M	SO	Business Administration	high	100	Converger	Concept Attainment	1
68	20	M	JR	Art	moderate	100	Accommodator	Direct Instruction	7
69	22	M	SR	Information Technology Leadership	high	300	Converger	Direct Instruction	7
70	22	F	SR	Business Administration	high	200	Accommodator	Direct Instruction	4
71	19	F	SO	Information Technology Leadership	very high	100	Converger	Direct Instruction	9
72	22	F	SR	Accounting	moderate	200	Accommodator	Concept Attainment	5
73	24	M	JR	Thematic Major Information Technology Leadership	moderate	100	Accommodator	Concept Attainment	6
74	20	M	SO	Information Technology Leadership	high	100	Accommodator	Concept Attainment	5
75	21	M	JR	English	very high	200	Converger	Direct Instruction	6



76	21	M	JR	Information Technology Leadership	high	200	Diverger	Concept Attainment	7
77	20	M	JR	Art Information Technology Leadership	moderate	100	Assimilator	Concept Attainment	4
78	21	F	SR	Information Technology Leadership	high	200	Converger	Concept Attainment Direct	10
79	22	F	SR	Accounting Business	moderate	200	Accomodator	Instruction Direct	3
80	22	M	SR	Administration Business	moderate	100	Converger	Instruction Direct	8
81	20	M	SO	Administration Information Technology Leadership	very high	200	Converger	Instruction	6
82	20	M	SO	Business Administration	very high	100	Assimilator	Concept Attainment Direct	8
83	20	M	SO	Administration	high	100	Converger	Instruction Direct	6
84	22	M	SR	Accounting Business	high	200	Assimilator	Instruction Concept	6
85	21	M	JR	Administration	very high	200	Converger	Attainment Concept	7
86	22	M	SR	Accounting	high	200	Accomodator	Attainment	7